

Lexikos 20

Lexikos 20

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 20:2010

AFRILEX SERIES 20:2010



BURO VAN DIE WAT

STELLENBOSCH

Die publikasie van hierdie boek is moontlik gemaak deur 'n ruim skenking van die L.W. Hiemstra-Trust – Opgerig deur Riekie Hiemstra ter herinnering aan Ludwig Wybren (Louis) Hiemstra.

The publication of this book was made possible by a generous donation from the L.W. Hiemstra Trust – Established by Riekie Hiemstra in memory of Ludwig Wybren (Louis) Hiemstra.

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2010 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2010

Tipografie en uitleg
deur Tanja Harteveld en Riette Ruthven
Bandontwerp deur Piet Grobler
Geset in 10 op 12 pt Palatino

Gedruk en gebind deur Onyx Press
Stewartstraat 61 Goodwood

ISBN 978-0-9814434-2-3
ISSN 1684-4904

Geen gedeelte van hierdie publikasie mag sonder skriftelike verlof van die uitgewer gereproduseer of in enige vorm of deur enige elektroniese of meganiese middel weergegee word nie, hetsy deur fotokopiëring, plaat- of bandopname, mikroverfilming of enige ander stelsel van inligtingsbewaring

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, including electronic, mechanical, photographic, magnetic or other means, without the prior written permission of the publisher

Menings wat in artikels en resensies uitgespreek word, is nie noodwendig dié van AFRILEX of die Buro van die WAT nie

Opinions expressed in the articles and reviews are not necessarily those of AFRILEX or of the Bureau of the WAT

Lexikos is elektronies beskikbaar by Sabinet Online en African Journals OnLine (AJOL)
Lexikos is available online from Sabinet Online and African Journals OnLine (AJOL)

Indekse Indexes

Arts and Humanities Citation Index®, Current Contents®/Arts & Humanities, Current Contents®/Social and Behavioral Sciences, Journal Citation Reports/Social Sciences Edition, Social Sciences Citation Index®, and Social Scisearch®; Linguistic Bibliography Online; Linguistics Abstracts Online; Linguistics and Language Behavior Abstracts; MLA International Bibliography; R.R.K. Hartmann's Bibliography of Lexicography; Scopus

Adviesraad / Advisory Board

- Prof. A. Adamska-Salaciak (Pole/Poland)
Prof. H. Béjoint (Frankryk/France)
Prof. H. Chimhundu (Zimbabwe)
Prof. F. Dolezal (VSA/USA)
Prof. R.H. Gouws (RSA)
Prof. R.R.K. Hartmann (Groot-Brittanje/Great Britain)
Prof. M.H. Heliel (Egipte/Egypt)
Dr. V. Kukanda (Gaboen/Gabon)
Prof. W. Martin (België en Nederland/Belgium and The Netherlands)
Prof. I.A. Mel'čuk (Kanada/Canada)
Prof. A.M.F.J. Moerdijk (Nederland/The Netherlands)
Prof. M. Schlaefer (Duitsland/Germany)
Dr. J. Tent (Australië/Australia)
Prof. J. Van Keymeulen (België/Belgium)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)

Redaksiekomitee / Editorial Committee

- Prof. H.M. Batibo (Botswana)
Dr. H.L. Beyer (Namibië/Namibia)
Prof. W.A.M. Carstens (RSA)
Mnr. E. Chabata (Zimbabwe)
Dr. A.R. Chuwa (Tanzanië/Tanzania)
Prof. C.J. Conradie (RSA)
Prof. J.D. Emejulu (Gaboen/Gabon)
Prof. A.E. Feinauer (RSA)
Prof. R. Finlayson (RSA)
Dr. S. Hadebe (Zimbabwe)
Prof. I.M. Kosch (RSA)
Dr. P.A. Louw (RSA)
Mnr. K.J. Mashamaite (RSA)
Dr. P.A. Mavoungou (Gaboen/Gabon)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Prof. B.M. Mini (RSA)
Dr. V.M. Mojela (RSA)
Mnr. M.C. Mphahlele (RSA)
Mnr. M.H. Mpungose (RSA)
Dr. H.S. Ndinga-Koumba-Binza (Gaboen/Gabon)
Mnr. D. Nkomo (Zimbabwe)
Prof. A.N. Otto (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

| | |
|-----------------------------------|------|
| Voorwoord | xi |
| Foreword | xiii |
| <i>J.C.M.D. du Plessis</i> | |
| 'n Woord van AFRILEX | xv |
| A Few Words from AFRILEX | xvii |
| <i>Gilles-Maurice de Schryver</i> | |
| Redaksionele doelstellings | xix |
| Editorial Objectives | xx |
| Redaktionelle Ziele | xxi |
| Politique éditoriale | xxii |

Artikels / Articles

| | |
|---|-----|
| The OEAD: New Perspectives on English–Arabic Dictionaries for English-Speaking Users and Users' Surveys | 1 |
| <i>Radia Benzehra and Don R. McCreary</i> | |
| A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography | 26 |
| <i>Henning Bergenholtz and Rufus H. Gouws</i> | |
| Kontekstuele bepalers by die beplanning van skoolwoordeboeke | 52 |
| <i>Herman L. Beyer</i> | |
| The Utilisation of Outer Texts in the Practical Lexicography of African Languages | 73 |
| <i>Emmanuel Chabata and Dion Nkomo</i> | |
| The Role of Dictionaries in the Documentation and Codification of African Languages: The Case of Khoisan | 92 |
| <i>Andy Chebanne</i> | |
| A Bilingualised English Dictionary for Catalan Speakers The EED | 109 |
| <i>Montse Corrius and Dídac Pujol</i> | |

| | |
|--|-----|
| Een digitaal compilatiecorpus historisch Nederlands <i>Evie Coussé</i> | 123 |
| Remarks on the Elaboration of an English–Spanish Word- combination Dictionary <i>Eduardo de Gregorio-Godeo</i> | 143 |
| Revolutionizing Bantu Lexicography — A Zulu Case Study <i>Gilles-Maurice de Schryver</i> | 161 |
| L'ingénierie lexicale ou la description d'un objet entre l'invention et la découverte <i>Jean-Nicolas De Surmont</i> | 202 |
| Le dictionnaire comme outil d'enseignement des langues au Gabon <i>Guy-Modeste Ekwa Ebanéga</i> | 212 |
| Does Johnson's Prescriptive Approach Still Have a Role to Play in Modern-Day Dictionaries? <i>Rufus H. Gouws and Liezl Potgieter</i> | 234 |
| Firming up the Foundations: Reflections on Verifying the Quotations in a Historical Dictionary, with Reference to <i>A Dictionary of South African English on Historical Principles</i> <i>Sheila Hicks</i> | 248 |
| Access Routes of Internet Finance Dictionaries: Present Solutions and Future Opportunities <i>Deny Arnos Kwary</i> | 272 |
| Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries <i>Robert Lew</i> | 290 |
| The <i>Tshivenda–English Ṭhalusamaipfi / Dictionary</i> as a Product of South African Lexicographic Processes <i>Mbulungeni Madiba and Dion Nkomo</i> | 307 |
| The Monolingual Lusoga Dictionary Faced with Demands from a New User Category <i>Minah Nabirye and Gilles-Maurice de Schryver</i> | 326 |

| | |
|--|-----|
| The Treatment of Polysemy and Homonymy in Monolingual General-purpose Dictionaries with Special Reference to <i>Isichazamazwi SesiNdebele</i> <i>Eventhough Ndlovu and Sanelisiwe Sayi</i> | 351 |
| Affirming a Role for Specialised Dictionaries in Indigenous African Languages <i>Dion Nkomo</i> | 371 |
| Die verifiëring, verfyning en toepassing van leksikografiese liniale vir Afrikaans <i>D.J. Prinsloo</i> | 390 |
| Checking Knowledge in Online Encyclopaedias: Towards a Behavioural Approach to Data Accessibility <i>Marcin Overgaard Ptaszynski</i> | 410 |
| Improving the Functionality of Dictionary Definitions for Lexical Sets: The Role of Definitional Templates, Definitional Consistency, Definitional Coherence and the Incorporation of Lexical Conceptual Models <i>Piet Swanepoel</i> | 425 |
| Reflections on the Academic Status of Lexicography <i>Sven Tarp</i> | 450 |
| Skoolwoordeboeke vir huistaalleerders van Afrikaans <i>Sven Tarp en Rufus H. Gouws</i> | 466 |
| Hybride textuelle Strukturen und hybride textuelle Einheiten. Ein Beitrag zur Theorie der Wörterbuchform <i>Herbert Ernst Wiegand</i> | 495 |
| Countable or Uncountable? That is the Question — Lexicographic Solutions to Nominal Countability in Learner's Dictionaries for Production Purposes <i>Mei Xue</i> | 540 |
| Oor die samestelling van eentalige woordeboeke / On the Compilation of Monolingual Dictionaries | |
| Notes on Compiling a Corpus-Based Dictionary <i>František Čermák</i> | 559 |

- Compiling a Monolingual Dictionary for Native Speakers 580
Patrick Hanks

Terminologiebestuur / Terminology Management

- National Language and Terminology Policies — A South African Perspective 599
Mariëtta Alberts
- Terminologiebestuur in Suid-Afrika: 'n Ideaalmodel 621
Dina Ferreira en Theodorus du Plessis

Leksiko-opname / Lexicosurvey

- Namibian University Entrants' Concepts of 'a Dictionary' 644
Herman L. Beyer and André K. Faul
- The Use of Pocket Electronic Dictionaries by Thai University Students 680
Atipat Boonmoh

Leksikonotas / Lexiconotes

- Borrowing and Dictionary Compilation: The Case of the Indigenous South African Languages 691
Munzhedzi James Mafela
- Borrowing and Loan Words: The Lemmatizing of Newly Acquired Lexical Items in Sesotho sa Leboa 700
V.M. Mojela
- The Effects of Shona Language Change on Monolingual Lexicography: The Need for a Revised Alphabet 708
William Zivenge, Gift Mheta and Maxwell Kadenge

Resensieartikels / Review Articles

- More than Twenty Years of Svensén's *Handbook* 716
Loránd-Levente Pálfi
- H.A. Welker and Pedagogical Lexicography 730
Sven Tarp

Resensies / Reviews

- W.F. Botha (Hoofredakteur): *Woordeboek van die Afrikaanse Taal, Dertiende Deel: R.* 750
J.D. McLachlan
- G.-M. de Schryver (Editor): *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks* 757
Robert Lew
- G.-M. de Schryver (Editor): *Oxford Bilingual School Dictionary Zulu and English* 760
D.J. Prinsloo
- Jean Nicolas De Surmont: *Chanson: son histoire et sa famille dans les dictionnaires de langue française* 767
François Jaques
- Elektroniese WAT (Woordeboek van die Afrikaanse Taal, A tot R)* 770
Corlietha Swart

Publikasieaankondigings / Publication Announcements

- Voorskrifte aan Skrywers 775
Instructions to Authors 776
Richtlinien für Autoren 778
Instructions aux Auteurs 779

Voorwoord

Vanjaar se nommer van *Lexikos* is 'n omvangryke bundel. Dit bevat 26 artikels wat 'n wye reeks leksikografiese onderwerpe vanuit verskeie hoeke benader en wat 'n verskeidenheid woordeboeke in verskillende tale bespreek. 'n Spesiale afdeling oor die samestelling van eentalige woordeboeke bied twee komplementêre artikels aan. In 'n verdere spesiale afdeling word Suid-Afrikaanse terminologiebestuur en -beleid in twee aanvullende artikels bespreek. Twee leksikografiese opnames oor uiteenlopende onderwerpe word beskryf: oor studente se opvattinge van woordeboeke en oor studente se gebruik van sakelektroniese woordeboeke. Die afdeling met leksikografiese notas bevat drie artikels oor die inkorporering en opname van leenwoorde in die Afrikatale en Afrikataalwoordeboeke. Twee resensieartikels en vyf resensies, gevolg deur twee bladsye publikasieaankondigings, sluit die inhoud af.

Hierdie nommer van *Lexikos* is die laaste wat ek sal redigeer. Nadat ek veertien jaar die redakteur van *Lexikos* was, het die tyd aangebreek om die redakteurskap aan iemand anders oor te dra. *Lexikos* sal nou op 'n roterende basis deur proff. E. Taljard, D.J. Prinsloo en R.H. Gouws geredigeer word. Ek wens hulle alles van die beste toe vir die komende jare.

Aan die begin van die jaar het prof. J. Taldeman *Lexikos* ingelig dat hy uit die Adviesraad wil tree. *Lexikos* was gelukkig om die afgelope sestien jaar uit sy kundigheid te kon voordeel trek. Ek wil my beste wense aan hom oordra vir die toekoms. Sy plek word gevul deur prof. J. Van Keymeulen wat nie onbekend is aan *Lexikos* nie: hy het sowel artikels vir publikasie bygedra as bydraers vir *Lexikos* gekeur. Ek hoop sy verbintenis met *Lexikos* sal vir baie jare voortduur.

Gedurende die afgelope jare het die Redaksiekomitee van *Lexikos* talle kere verander, maar verskeie kernlede het gebly sedert die eerste Redaksiekomitee in 1994 saamgestel is. Dit is daarom aangenaam om twee nuwe lede te verwelkom: prof. A.N. Otto en mnr. D. Nkomo. Hulle is gereelde bydraers tot en keurders vir *Lexikos*. Hopelik sal hulle voortgaan om 'n vrugbare verbintenis met *Lexikos* te hê.

Elke nommer van *Lexikos* is 'n spanpoging. Die bydraers is belangrike lede van die span. Ewe belangrik egter is die keurders. Alhoewel baie artikels in 'n afgeronde vorm ontvang word, word baie ander aansienlik verbeter deur opmerkings en voorstelle van die keurders. Vir my as redakteur was dit altyd die bevredigendste en lonendste ondervinding om te sien hoe artikels op hierdie manier ontwikkel. Ek wil daarom al die keurders bedank wat gedurende die afgelope jare bygedra het om die gehalte van baie van die artikels in *Lexikos* te verbeter.

Die laaste tien jaar, sedert 2001, was me. Riette Ruthven verantwoordelik vir die setwerk van *Lexikos*. Behalwe haar vaardigheid op die rekenaar, het sy

Voorwoord

'n oog vir besonderhede en uitleg wat sowel die gehalte as die voorkoms van *Lexikos* verbeter het. Ek wil haar bedank vir haar aandeel in die totstandbrenging van *Lexikos* wat ongetwyfeld bygedra het tot sy toenemende sukses. Sy was ook grotendeels verantwoordelik vir die uitgebreide administratiewe pligte wat die las van die redakteur aansienlik verlig het. Die setwerk van toekomstige nommers sal hanteer word deur me. Hermien van der Westhuizen wat saam met me. Tanja Harteveld, me. Hanlie Meitzler en mnr. Etienne Botha die setwerk van die eerste agt nommers van *Lexikos* behartig het. Dit is aangenaam om haar terug te verwelkom. Hopelik sal haar hernude verbintenis met *Lexikos* net so stimulerend soos vroeër wees.

Oor die jare het me. Tanja Harteveld, behalwe vir haar bydrae tot die setwerk, gehelp met die oplossing van talle rekenaarhaakplekke en -probleme. Sy sal in hierdie hoedanigheid voortgaan. Behalwe dat sy as Resensieredakteur sal optree, sal sy ook verantwoordelik wees vir die administratiewe pligte van *Lexikos*. Ek wil haar bedank vir haar hulp in die verlede en wens haar 'n voortgaande suksesvolle verbintenis met *Lexikos* in die toekoms.

J.C.M.D. du Plessis
Buro van die Woordeboek van die Afrikaanse Taal

Foreword

This year's issue of *Lexikos* is a comprehensive volume. It contains 26 articles approaching a wide range of lexicographic topics from several angles and discussing a variety of dictionaries in different languages. A special section on the compilation of monolingual dictionaries presents two complementary articles. In a further special section, South African terminology management and policy are discussed in two supplementary articles. Two lexicographic surveys on diverse subjects are described: on students' concepts of dictionaries and on students' use of pocket electronic dictionaries. The section with lexicographic notes contains three articles on the incorporation and inclusion of loan words in the African languages and African language dictionaries. Two review articles and five reviews, followed by two pages of publication announcements, conclude the contents.

This issue of *Lexikos* is the last that I will edit. After having been the editor of *Lexikos* for fourteen years, the time has come to pass on the editorship to somebody else. *Lexikos* will now be edited on a rotating basis by Proff. E. Taljard, D.J. Prinsloo and R.H. Gouws. I wish them everything of the best for the coming years.

At the beginning of the year Prof. J. Taeldeman informed *Lexikos* that he wants to resign from the Advisory Board. *Lexikos* has been fortunate that it could benefit from his expertise for the past sixteen years. I want to extend my best wishes to him for the future. His place has been filled by Prof. J. Van Keymeulen who is not unknown to *Lexikos*: he has both contributed articles for publication and refereed contributions to *Lexikos*. I hope his association with *Lexikos* will continue for many years.

During the past years, the Editorial Committee of *Lexikos* has changed many times, but several core members have remained since the first Editorial Board was composed in 1994. It is therefore a pleasure to welcome two new members: Prof. A.N. Otto and Mr. D. Nkomo. They are regular contributors to and referees for *Lexikos*. Hopefully they will continue to have a fruitful association with *Lexikos*.

Every issue of *Lexikos* is a team effort. The contributors are important members of the team. However, equally important are the referees. Although many articles are received in an accomplished form, many others have been considerably improved by comments and suggestions of the referees. For me as editor, it has always been a most satisfactory and rewarding experience to see how articles developed in this way. I therefore want to thank the referees who have helped during the past years to enhance the quality of the articles in *Lexikos*.

For the past ten years, since 2001, Ms Riette Ruthven has been responsible for the typesetting of *Lexikos*. Apart from her efficiency on the computer, she

Foreword

has an eye for detail and layout which has improved both the quality and appearance of *Lexikos*. I want to thank her for her part in the production of *Lexikos* which was undoubtedly instrumental in its increasing success. She was also largely responsible for the extensive administrative duties of *Lexikos* which considerably lightened the burden of the editor. The typesetting of future issues will be handled by Ms Hermien van der Westhuizen, who together with Ms Tanja Harteveld, Ms Hanlie Meitzler and Mr Etienne Botha managed the typesetting of the first eight issues of *Lexikos*. It is pleasant to welcome her back. Hopefully her renewed association with *Lexikos* will prove as stimulating as before.

Over the years, Ms Tanja Harteveld has, apart from her contribution to the typesetting, assisted in solving many computational hitches and problems. In addition to her acting as Review Editor, she will also be responsible for the administrative duties of *Lexikos*. I want to thank her for her help in the past and wish her a continuing successful connection with *Lexikos* in the future.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal

'n Woord van AFRILEX

'n Totaal van vyf kontinentale verenigings vir die leksikografie is tans aktief. Alhoewel hierdie verdeling heeltemal kunsmatig is — omdat elke vereniging leksikografiese bydraes in enigeen van die wêreld se tale verwelkom — bly die skeiding prakties. AFRILEX, gestig in 1995, is een van die jonger familieledes, aangesien die DSNA reeds twintig jaar vroeër tot stand gekom het. Drie van die vyf kontinentale verenigings het ook "hulle" eie tydskrif soos gesien kan word in die tabel hieronder.

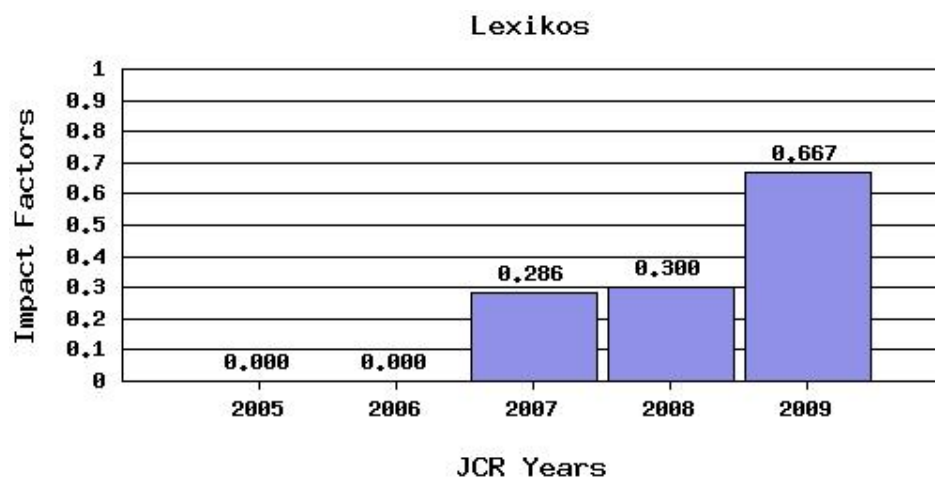
| Vereniging | Akroniem | Gestig | Tydskrif | Sedert |
|---|-----------------|---------------|---------------------|---------------|
| The Dictionary Society of North America | DSNA | 1975 | <i>Dictionaries</i> | 1979 |
| European Association for Lexicography | EURALEX | 1983 | IJL | 1988 |
| Australasian Association for Lexicography | AUSTRALEX | 1990 | — | — |
| African Association for Lexicography | AFRILEX | 1995 | <i>Lexikos</i> | 1991 |
| Asian Association for Lexicography | ASIALEX | 1997 | — | — |

Die eerste nommer van die tydskrif *Dictionaries*, die lyfblad van die DSNA, het vier jaar na die vereniging se totstandkoming verskyn. Bydraers tot *Dictionaries* moet lede van die DSNA in die publikasiejaar van hulle bydraes wees. Die eerste nommer van die *International Journal of Lexicography* (IJL), die lyfblad van EURALEX, het vyf jaar na die vereniging se totstandkoming verskyn. Bydraes van enige vakkundiges word vir IJL aanvaar. Die verwantskap tussen AFRILEX en *Lexikos* is egter effens anders, omdat die eerste nommer van "ons" tydskrif vier jaar *voor* die totstandkoming van ons vereniging verskyn het. In werklikheid het die inisiatief om die tydskrif *Lexikos* van stapel te stuur van die Buro van die *Woordeboek van die Afrikaanse Taal* (WAT) gekom, en toe AFRILEX gestig is, was ons vereniging gelukkig genoeg om deur 'n reeds suksesvolle publikasie gedra te word. Deur die jare het die redakteur van *Lexikos* die voorkeure begin verwag op hersiene referate wat aanvanklik gelewer is by AFRILEX-konferensies, en bykomend bydraes verwelkom uit enige deel van die wêreld oor enige taal (tale). Wat aldrie tydskrifte — *Dictionaries*, IJL, en *Lexikos* — in gemeen het, is dat hulle ingesluit is by die lidmaatskap van die onderskeie verenigings — die DSNA, EURALEX en AFRILEX. Ledegeld vir AFRILEX dek egter nie die koste om *Lexikos* te publiseer nie, sodat ons steeds erkenning moet verleen aan die Buro van die WAT vir sy vermoë om donasies te verkry om hierdie wonderlike publikasie te subsidieer. Wanneer daar dan na

'n Woord van AFRILEX

die volle prentjie gekyk word, was AFRILEX werklik baie gelukkig om op soveel vlakke met die Buro saam te werk, en dit reeds vir vyftien jaar.

Hierdie goeie gelukslae het nie daar geëindig nie. Soos ek in my Woord van AFRILEX in verlede jaar se nommer daarop gewys het, is *Lexikos* nou 'n geïndekseerde tydskrif waarvoor 'n impakfaktor bereken word in die kategorie Linguistiek. Soos gesien kan word uit die grafiek hieronder, het die laaste (2009) impakfaktor van *Lexikos* meer as verdubbel in vergelyking met die vorige een.



Om die geheue te verfris, die impakfaktor word soos volg bereken: Vir enige besondere jaar word die aantal verwysings na artikels gepubliseer in die twee vorige jare verdeel deur die totale aantal artikels gepubliseer in hierdie twee jaar. In die 2009 Journal Citation Reports gee dit die volgende vir *Lexikos*:

| Verwysings in 2009 na artikels gepubliseer in: | Aantal artikels gepubliseer in: |
|--|---------------------------------|
| 2008 = 14 | 2008 = 21 |
| 2007 = 20 | 2007 = 30 |
| Som: 34 | Som: 51 |

$$\text{Impakfaktor} = \frac{\text{Verwysings na onlangse artikels}}{\text{Aantal onlangse artikels}} = \frac{34}{51} = 0.667$$

Met 'n impakfaktor van 0.667, is *Lexikos* nou 49ste in rang uit 92 tydskrifte in die Linguistiekategorie, wat dit in die sogenaamde derde kwartiel plaas (op vanaf die vierde kwartiel). Met die impak van *Lexikos* nou dubbel van wat dit was, ons hartlike dank aan sy redakteur, dr. J.C.M.D. du Plessis, sowel as sy regterhand, me. Riette Ruthven, gelyklopend verdubbel!

Gilles-Maurice de Schryver
President: AFRILEX

A Few Words from AFRILEX

A total of five continental associations for lexicography are currently active. Although this division is entirely artificial — as each association welcomes lexicographic endeavours in any of the world's languages — the cut-up remains practical. AFRILEX, founded in 1995, is one of the younger siblings, given that the DSNAs had already been created twenty years earlier. Three of the five continental associations also have 'their' own journal, as may be seen in the table below.

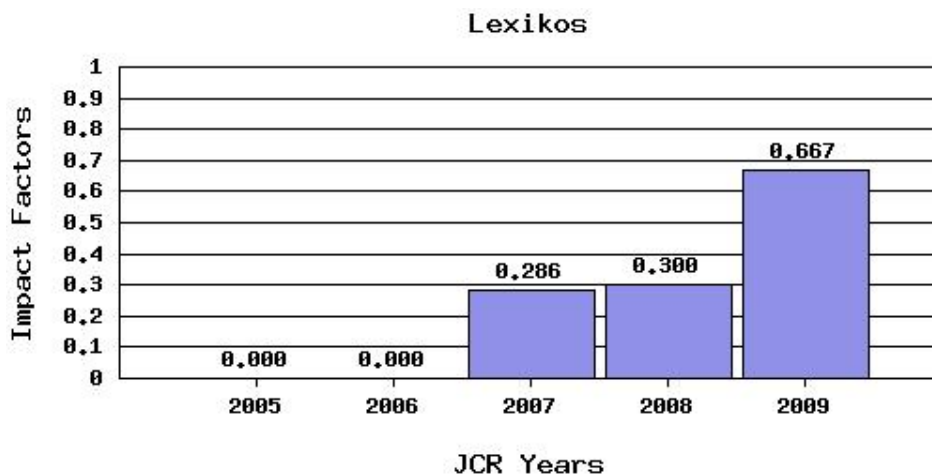
| Association | Acronym | Founded | Journal | Since |
|---|-----------|---------|---------------------|-------|
| The Dictionary Society of North America | DSNA | 1975 | <i>Dictionaries</i> | 1979 |
| European Association for Lexicography | EURALEX | 1983 | IJL | 1988 |
| Australasian Association for Lexicography | AUSTRALEX | 1990 | — | — |
| African Association for Lexicography | AFRILEX | 1995 | <i>Lexikos</i> | 1991 |
| Asian Association for Lexicography | ASIALEX | 1997 | — | — |

The first volume of the journal *Dictionaries*, the mouthpiece of the DSNAs, appeared four years after that association's establishment. Contributors to *Dictionaries* must be members of the DSNAs in the year of publication of their contributions. The first issue of the *International Journal of Lexicography* (IJL), the mouthpiece of EURALEX, appeared five years after that association's establishment. Contributions from any scholars are accepted in IJL. The relationship between AFRILEX and *Lexikos*, however, is slightly different, as the first volume of 'our' journal appeared four years *before* the establishment of our association. In effect, the initiative to launch the journal *Lexikos* came from the Bureau of the *Woordeboek van die Afrikaanse Taal* (WAT), and when AFRILEX was established, our association was lucky enough to be able to piggyback on an already successful publication. Over the years, the editor of *Lexikos* has come to expect a right of first refusal for reworked papers initially read at AFRILEX conferences, and additionally welcomes material from any part of the world, on any language(s). What all three journals — *Dictionaries*, IJL, and *Lexikos* — have in common is that they are included with a membership of the respective associations — the DSNAs, EURALEX, and AFRILEX. Membership fees for AFRILEX, however, do not cover the cost to produce *Lexikos*, so we should continue to be grateful to the Bureau of the WAT for its ability to attract dona-

A Few Words from AFRILEX

tions with which to subsidize this wonderful publication. Looking at the full picture, then, AFRILEX has been very fortunate indeed to be able to cooperate with the Bureau on so many levels, and this for fifteen years already.

These strokes of good fortune did not end there. As I pointed out in my Word from AFRILEX in last year's volume, *Lexikos* is now an indexed journal for which an Impact Factor is being calculated, in the category Linguistics. As may be seen from the graph below, the latest (2009) Impact Factor of *Lexikos* has more than doubled compared to the previous one:



As a reminder, here is how Impact Factors are calculated: For any particular year, one divides the number of cites to articles published in the two previous years, by the total number of articles published in those two years. In the 2009 Journal Citation Reports, this gives the following for *Lexikos*:

| Cites in 2009 to articles published in: | Number of articles published in: |
|---|----------------------------------|
| 2008 = 14 | 2008 = 21 |
| 2007 = 20 | 2007 = 30 |
| Sum: 34 | Sum: 51 |

$$\text{Impact Factor} = \frac{\text{Cites to recent articles}}{\text{Number of recent articles}} = \frac{34}{51} = 0.667$$

With an Impact Factor of 0.667, *Lexikos* now ranks 49th out of 92 journals in the Linguistics category, which places it in the so-called third quartile (up from the fourth quartile). With the impact of *Lexikos* now double of what it used to be, our heartfelt thanks to its editor, Dr J.C.M.D. du Plessis, as well as his right hand, Ms Riette Ruthven, redouble concurrently!

Gilles-Maurice de Schryver
President: AFRILEX

Redaksionele doelstellings

Lexikos is 'n tydskrif vir die leksikografiese vakspesialis en word in die AFRILEX-reeks uitgegee. "AFRILEX" is 'n akroniem vir "leksikografie in en vir Afrika". Van die sesde uitgawe af dien *Lexikos* as die amptelike mondstuk van die *African Association for Lexicography* (AFRILEX), onder meer omdat die Buro van die WAT juis die uitgesproke doel met die uitgee van die AFRILEX-reeks gehad het om die stigting van so 'n leksikografiese vereniging vir Afrika te bevorder.

Die strewe van die AFRILEX-reeks is:

- (1) om 'n kommunikasiekanaal vir die nasionale en internasionale leksikografiese gesprek te skep, en in die besonder die leksikografie in Afrika met sy ryk taleverskeidenheid te dien;
- (2) om die gesprek tussen leksikograwe onderling en tussen leksikograwe en taalkundiges te stimuleer;
- (3) om kontak met plaaslike en buitelandse leksikografiese projekte te bewerkstellig en te bevorder;
- (4) om die interdisiplinêre aard van die leksikografie, wat ook terreine soos die taalkunde, algemene taalwetenskap, leksikologie, rekenaarwetenskap, bestuurskunde, e.d. betrek, onder die algemene aandag te bring;
- (5) om beter samewerking op alle terreine van die leksikografie moontlik te maak en te koördineer, en
- (6) om die doelstellings van die *African Association for Lexicography* (AFRILEX) te bevorder.

Hierdie strewe van die AFRILEX-reeks sal deur die volgende gedien word:

- (1) Bydraes tot die leksikografiese gesprek word in die vaktydskrif *Lexikos* in die AFRILEX-reeks gepubliseer.
- (2) Monografiese en ander studies op hierdie terrein verskyn as afsonderlike publikasies in die AFRILEX-reeks.
- (3) Slegs bydraes wat streng vakgerig is en wat oor die suiwer leksikografie of die raakvlak tussen die leksikografie en ander verwante terreine handel, sal vir opname in die AFRILEX-reeks kwalifiseer.
- (4) Die wetenskaplike standaard van die bydraes sal gewaarborg word deur hulle aan 'n komitee van vakspesialiste van hoë akademiese aansien voor te lê vir anonieme keuring.

Lexikos sal jaarliks verskyn, terwyl verdienstelike monografiese studies sporadies en onder hulle eie titels in die AFRILEX-reeks uitgegee sal word.

Editorial Objectives

Lexikos is a journal for the lexicographic specialist and is published in the AFRILEX Series. "AFRILEX" is an acronym for "lexicography in and for Africa". From the sixth issue, *Lexikos* serves as the official mouthpiece of the *African Association for Lexicography* (AFRILEX), amongst other reasons because the Bureau of the WAT had the express aim of promoting the establishment of such a lexicographic association for Africa with the publication of the AFRILEX Series.

The objectives of the AFRILEX Series are:

- (1) to create a vehicle for national and international discussion of lexicography, and in particular to serve lexicography in Africa with its rich variety of languages;
- (2) to stimulate discourse between lexicographers as well as between lexicographers and linguists;
- (3) to establish and promote contact with local and foreign lexicographic projects;
- (4) to focus general attention on the interdisciplinary nature of lexicography, which also involves fields such as linguistics, general linguistics, lexicology, computer science, management, etc.;
- (5) to further and coordinate cooperation in all fields of lexicography; and
- (6) to promote the aims of the *African Association for Lexicography* (AFRILEX).

These objectives of the AFRILEX Series will be served by the following:

- (1) Contributions to the lexicographic discussion will be published in the specialist journal *Lexikos* in the AFRILEX Series.
- (2) Monographic and other studies in this field will appear as separate publications in the AFRILEX Series.
- (3) Only subject-related contributions will qualify for publication in the AFRILEX Series. They can deal with pure lexicography or with the intersection between lexicography and other related fields.
- (4) Contributions are judged anonymously by a panel of highly-rated experts to guarantee their academic standard.

Lexikos will be published annually, but meritorious monographic studies will appear as separate publications in the AFRILEX Series.

Redaktionelle Ziele

Lexikos ist eine Zeitschrift für Fachleute der Lexikographie, die in der AFRILEX-Serie erscheint. "AFRILEX" ist ein Akronym für "Lexikographie in und für Afrika". Von der sechsten Ausgabe an dient *Lexikos* als amtliches Sprachrohr der *African Association for Lexicography* (AFRILEX), u.a. weil das Büro des WAT das gerade angesprochene Ziel mit der Ausgabe der AFRILEX-Serie verfolgt, die Gründungsziele eines solchen lexikographischen Vereins für Afrika zu fördern.

Die folgenden Ziele werden mit den Publikationen der AFRILEX-Serie verfolgt: Man möchte:

- (1) ein Medium schaffen für die nationale und internationale Diskussion, besonders aber der Lexikographie in Afrika mit seinen zahlreichen Sprachen dienen;
- (2) die Diskussion fördern, unter Lexikographen als auch zwischen Lexikographen und Linguisten;
- (3) Kontakt herstellen und fördern zwischen südafrikanischen und ausländischen lexikographischen Projekten;
- (4) die Aufmerksamkeit lenken auf die interdisziplinäre wissenschaftliche Praxis der Lexikographie, die Beziehung aufweist zur Linguistik, allgemeinen Sprachwissenschaft, Lexikologie, Computerwissenschaft, zum Management und zu anderen Bereichen;
- (5) die Zusammenarbeit auf allen Gebieten der Lexikographie fördern und koordinieren;
- (6) die Ziele der *African Association for Lexicography* (AFRILEX) fördern.

Gemäß den Zielsetzungen der AFRILEX-Serie werden:

- (1) Beiträge zur lexikographischen Diskussion in der Fachzeitschrift *Lexikos* veröffentlicht;
- (2) monographische und andere Studien auf diesem Gebiet als getrennte Publikationen in der AFRILEX-Serie erscheinen;
- (3) nur einschlägige Beiträge, die sich ausschließlich mit Lexikographie oder mit fachverwandten Gebieten befassen, für Aufnahme in der AFRILEX-Serie in Betracht gezogen;
- (4) Beiträge anonym von einem aus Spezialisten des Faches von hohem akademischen Ansehen bestehenden Ausschuß beurteilt.

Lexikos erscheint jährlich. Ausgewählte monographische Studien dagegen erscheinen gelegentlich als getrennte Publikationen in der AFRILEX-Serie.

Politique éditoriale

La revue *Lexikos*, destinée aux spécialistes de lexicographie, est publiée dans la collection AFRILEX (acronyme de "lexicographie en Afrique et pour l'Afrique"). Depuis son sixième numéro, *Lexikos* est l'organe officiel de l'*African Association for Lexicography* (AFRILEX), entre autres parce que le Bureau du WAT s'est donné pour objectif de promouvoir le développement d'une telle association lexicographique en Afrique par la publication de la collection AFRILEX.

Les objectifs de la collection AFRILEX sont de :

- (1) créer un forum de discussion national et international sur la lexicographie, particulièrement au service de la lexicographie en Afrique, qui représente une grande diversité de langues;
- (2) stimuler le débat entre lexicographes, ainsi qu'entre lexicographes et linguistes;
- (3) établir et promouvoir le contact avec des projets lexicographiques locaux ou étrangers;
- (4) attirer l'attention générale sur la nature interdisciplinaire de la lexicographie, qui touche des domaines comme la linguistique générale, la lexicologie, l'informatique, le management, etc.;
- (5) favoriser et coordonner la coopération dans tous les domaines de la lexicographie; et
- (6) promouvoir les orientations de l'*African Association for Lexicography* (AFRILEX).

Pour atteindre ces objectifs, la collection AFRILEX

- (1) publiera les contributions aux discussions sur la lexicographie dans la revue *Lexikos*, dans la collection AFRILEX;
- (2) publiera sous forme de publications séparées dans la collection AFRILEX des monographies et autres travaux dans le domaine de la lexicographie;
- (3) ne publiera dans la série AFRILEX que des travaux dans le domaine de la lexicographie, qu'ils traitent de lexicographie pure ou des rapports entre la lexicographie et d'autres disciplines voisines; et
- (4) soumettra de manière anonyme toutes les propositions à des experts hautement qualifiés, pour en garantir le niveau académique.

Lexikos est publié annuellement, mais les travaux de qualité exceptionnelle seront publiés sous forme de publications séparées dans la collection AFRILEX.

The OEAD: New Perspectives on English–Arabic Dictionaries for English-Speaking Users and Users' Surveys

Radia Benzehra (*radia@uga.edu*)
and

Don R. McCreary (*mccreary@uga.edu*) *Linguistics Program,
University of Georgia, Athens, Georgia, United States of America*

The time is now ripe for a revised and enlarged edition of the OEAD, to maintain its position as a fitting partner to Wehr.

Haywood (1991: 3092)

Abstract: The article sheds light on the necessity of finding new techniques to rank the users' preferences for English–Arabic dictionaries. The first section of the article reports the findings of an experimental technique devised for this particular purpose. The reviews for dictionaries on Amazon.com turn out to be a more valuable source of lexicographical information than had been expected. It will also be shown how the reviews can determine the future buyers' choice as to which English–Arabic or Arabic–English dictionary would meet their needs. Based on the Amazon reviews, the article devotes a section to investigate the microstructural features of some lexical entries in the *Oxford English–Arabic Dictionary of Current Usage* (OEAD). To learn more about the dictionary's microstructure, the next section analyzes a pre-selected list of 113 lexical units in an attempt to test the dictionary for the amount of information it provides for some high-frequency items. The article deals with the OEAD mostly from the point of view of encoding English-speaking users. The Modern Language Association (MLA) recently reported that from 1998 to 2002 there was a 92% increase in the number of Arabic programs throughout the United States, hence the pressing need for dictionaries designed with English-speaking users as the target group. The article suggests that an extensive linguistic revision of the OEAD will make it more systematic and user-friendly.

Keywords: BILINGUAL DICTIONARY, DECODING USERS, ENCODING USERS, ENGLISH–ARABIC, ENGLISH-SPEAKING USERS, LEARNERS OF ARABIC, LEXICOGRAPHY, MICROSTRUCTURE, REVIEWS, USERS' PREFERENCES

Opsomming: Die OEAD: Nuwe perspektiewe op Engels–Arabiese woordeboeke vir Engelssprekende gebruikers en gebruikersopnames. Hierdie artikel werp lig op die noodsaaklikheid om nuwe tegnieke te kry om gebruikers se voorkeure vir Engels–Arabiese woordeboeke in rangorde te plaas. Die eerste afdeling van die artikel doen verslag van die bevindings van 'n eksperimentele tegniek ontwerp vir hierdie bepaalde doel. Die resensies van woordeboeke op Amazon.com het geblyk 'n meer waardevolle bron van leksikografiese inlig-

ting te wees as wat verwag is. Daar sal ook getoon word hoe die resensies toekomstige kopers se keuse kan bepaal oor watter Engels–Arabiese of Arabiese–Engelse woordeboek aan hul behoeftes sal voldoen. Met die Amazon-resensies as uitgangspunt, wy die artikel 'n afdeling aan die ondersoek van die mikrostrukturele eienskappe van sommige leksikale inskrywings in die *Oxford English–Arabic Dictionary of Current Usage* (OEAD). Om meer oor die woordeboek se mikrostruktuur te wete te kom, ontleed die volgende afdeling 'n voorafgekose lys van 113 leksikale eenhede in 'n poging om die woordeboek te toets vir die hoeveelheid inligting wat dit verskaf wat betref sommige hoëfrekwensie-items. Die artikel handel meestal oor die OEAD uit die gesigspunt van enkoderende Engelssprekende gebruikers. Die Modern Language Association het onlangs berig dat daar vanaf 1998 tot 2002 'n 92%-toename in die aantal Arabiese programme dwarsoor die Verenigde State was, vandaar die dringende behoefte aan woordeboeke vir Engelssprekende gebruikers as teikengroep. Die artikel suggereer dat 'n taalkundige hersiening van die OEAD dit meer sistematies en gebruikersvriendelik sal maak.

Slutelwoorde: TWEETALIGE WOORDEBOEK, DEKODERENDE GEBRUIKERS, ENKODERENDE GEBRUIKERS, ENGELS–ARABIES, ENGELSSPREKENDE GEBRUIKERS, AANLEERDERS VAN ARABIES, LEKSIKOGRAFIE, MIKROSTRUKTUUR, RESENSIES, GEBRUIKERSVOORKEURE

1. Introduction

The 1972 edition of the *Oxford English–Arabic Dictionary of Current Usage* (henceforth OEAD) is the dictionary recommended by teachers of Arabic at the University of Georgia, though they themselves admitted that using the dictionary was not a common practice in their classes. To highlight the underlying features of the OEAD, we examined this 1972 edition which has served English-speaking users for 37 years now. Because this edition has never been revised, we examined a few pages from its 1982 abridged version: *The Concise Oxford English–Arabic Dictionary of Current Usage* (henceforth COEAD). We noted two main differences which indicate that the 1972 edition is more useful for encoding English-speaking users than the abridged edition. The article aims at demonstrating how the OEAD (1972) can constitute a comparable alternative to the other bilingual dictionaries on the market. We do believe that it is time to bring the OEAD up to date with a specific category of users in mind.

The 2002 statistics on enrollments in foreign languages in U.S. institutions of higher education show that "enrollments in Arabic were relatively stable during the 1980s; however, since 1995 they have shown rapid growth, particularly between 1998 and 2002, almost doubling (from 5 505 to 10 584)" (Welles 2004: 14). Arabic is among the fifteen most commonly taught languages in the U.S., hence the pressing need for dictionaries designed for English-speaking users as the target group. But what does this category of users think of the different English–Arabic dictionaries available on the market? Which dictionaries do English-speaking learners of Arabic own or use on a regular basis? And what motivates their choice?

It is important to note that surveys of dictionary users' preferences are still difficult to conduct in a systematic way in American universities. In our university, for example, most learners of Arabic are beginners and are not familiar with English–Arabic dictionaries. They seem to be more absorbed in the intricacies of Arabic grammar than in dictionary use. The reasons behind the non-use of dictionaries in Arabic classrooms are not within the scope of this article. We are more concerned with finding out alternative techniques that can be used to rank the users' preferences for English–Arabic dictionaries in the United States. Such a ranking would enable us to understand the strengths and weaknesses of English–Arabic dictionaries from actual users' perspectives.

A potentially rewarding perspective on users' preferences is to consult the reviews for dictionaries on Amazon.com, which are numerous. Other websites for booksellers, such as Borders.com and Barnesandnoble.com, have a few reviews, but not as many as Amazon. This might be due to the fact that Amazon.com is one of the most popular websites for booksellers in the US, especially among students. Not only did the customers write reviews for the dictionaries they bought on Amazon, but also future buyers appear to take time and read the reviews posted online when trying to make a decision about buying a dictionary. The readers also leave comments as to whether or not the reviews have been helpful. The reviews take the form of a debate on different aspects of the dictionaries with eye-catching titles such as "Throw your Oxford in the Trash", and replies such as "Don't listen to the idiot who gave one star". The reviews are rated in terms of stars (5, 4, 3, 2 or 1 star) and each reviewer justifies the number of stars he assigns to the dictionary.

We should admit that the technique we present in the article still has shortcomings especially from a statistical point of view, as shall be explained later in the discussion, but we do believe that it provides lexicographers with valuable information on what users think of the different bilingual English–Arabic dictionaries they actually buy and use. The technique is also interesting in the sense that it provided us with a set of criteria along which we evaluated the OEAD from a dictionary user's perspective. In the last section of the article, we examined a pre-selected list of 113 lexical items with the aim of highlighting those areas where the OEAD or any other bilingual English–Arabic dictionary can be improved to — finally — meet the needs of a category of users so far neglected.

2. Survey of the users' preferences on Amazon.com

While the editor Doniach (1972) claims that the COEAD contains 35 000 words and phrases, he does not suggest any headword count in the front pages of the OEAD. Asfour (2003) obtained his own count of headwords for the dictionaries he analyzed by averaging the headwords in ten randomly selected pages and multiplying the average by the number of pages in the dictionary. We followed the same method and our own count of headwords in the OEAD is 26 726

entries.¹ The OEAD editor claims in his preface that the dictionary "is designed to meet the needs of those whose mother-tongue is English and who are learning Arabic, and those whose mother-tongue is Arabic and who are learning English" (1972: i). By no means can this ideal be achieved. Haas (1962) suggests that there should be a warning to the potential users that such is not actually the case. One can argue that there are innumerable facts about Arabic which are known to the native speaker, but which are not clear or obvious to the native speaker of English. Thus, a single dictionary cannot adequately help both Arabic speakers of English and English speakers of Arabic. The article focuses mainly on those linguistic areas which are reported by the Amazon reviewers as being problematic.

Considering the customer reviews on Amazon.com, the linguistic areas which seem to pose a problem for the users are mainly phonological and semantic. The users note that the Arabic vowels are sometimes suppressed from the written form in the dictionary. Short vowels in Arabic are indicated by means of diacritics, for example, a short oblique stroke written below the consonant indicates that the consonant is followed by a short /i/ (for more on Arabic diacritics see Schulz 2004). Our analysis indicates a lack of consistency in indicating the Arabic vowels, although one English–Arabic/Arabic–English dictionary by Hippocrene has attempted to solve this problem by using the IPA. What reviewers liked the most about the Hippocrene *Arabic Practical Dictionary* (2004) is the IPA system adopted in the A–E section. Each Arabic equivalent is transcribed in English; a strategy which facilitates the reading of Arabic words. About this pocket dictionary, one Amazon reviewer writes, "I can quickly find the words and the correct pronunciation. I would recommend it to anyone wanting to learn the Arabic language."

Also most users in the Amazon reviews accuse the two editions of the OEAD of not being accurate because when they use certain equivalents offered by both dictionaries, native speakers of Arabic tell them that they never use such words as can be illustrated from the following quote: "I often found natives saying "we never use this" when I would ask them about translations found in the [OEAD] dictionary. Very rarely, if ever, are the translations within it correct." The users report they get frustrated when they try to find some translations for entries that are either missing or even worse: when the English word has an entry but has multiple Arabic equivalents, no help is offered about which one would apply in a given context. One Amazon customer complains, "Sometimes there are multiple Arabic words for one English entry [in the OEAD]. I would find myself using the first word only, and then asking native speakers or my teacher if that was the right word." It should be noted, however, that the dictionary has been rated by some users as more English-speaker-friendly compared to other dictionaries. One review says, "True, the OEAD is not for beginners, but for anyone who can read Arabic, it is one of the best English–Arabic dictionaries you'll find." Another customer adds, "This is the only English to Arabic dictionary you'll find that gives you thorough examples

of usage. There is simply no substitute for the English to Arabic for Arabic students at any level."

About the COEAD (1982), a concise edition of the English–Arabic dictionary, one Amazon reviewer, among others, explains: "sometimes there are multiple Arabic words for one English entry. No help is offered about which one would apply to a given situation ...," which is the same criticism addressed to the OEAD. Let us consider the entry for **abandon** by way of illustration. The entry provides the COEAD user with five equivalents without any further indications as to the possible contexts where these words might occur: هَجَرَ /hajara/, تَرَكَ /taraka/, نَحَى تَخْلَى /takhala: ʕan/, نَحَى عَدَلَ /ʕadala ʕan/ and أَهْمَلَ /ahmala/. One sense discriminator is provided for the last equivalent (مَهَامَةٌ) /maha:mahu/ which translates literally as 'duties'. Sense discriminators written in Arabic are not always helpful for English-speaking users. Even in the OEAD, many Arabic equivalents appear in addition to sense discriminators written in Arabic. Such a practice is undoubtedly redundant for Arab users and suits more advanced learners of Arabic.

Another reviewer comments about the COEAD: "you will need a magnifying glass just to attempt to read the script." Indeed, when examining the abridged edition of the OEAD, we noted that the selected typeface is so small that the letters are difficult to read, especially for a new learner of Arabic. We also noted that the editor made some important decisions, such as the addition of the phonetic transcription of the English words, on the one hand, and the omission of many Arabic equivalents, on the other hand. A native speaker of English learning Arabic can dispense with any phonetic information for English; he will rather want more information included in the Arabic side. It might be argued that this information can be helpful for an Arab learner of English, though we do not see why a learner of English would consult a bilingual dictionary to check the pronunciation of an English word. Al-Kasimi (1977: 36) argued: "one can rightly imagine that information about pronunciation is more important for the foreign learner of the language. If he uses his bilingual dictionary to produce the foreign language he certainly wants to know the appropriate word and how to pronounce it."

The need for more information, especially for encoding users, is clearly felt because English and Arabic are so different at all linguistic levels: phonological, morphological, syntactic and cultural above all else. Morphology alone is one of the greatest challenging areas for learners of Arabic. Haywood (1991: 3089) explains that "Arabic has so complex an array of plural forms ('broken plurals') that it is essential for non-Arabs to be given the plurals of a large proportion of nouns and adjectives". Learners of Arabic need to use monolingual Arabic dictionaries which always give plural forms of nouns and adjectives unlike the bilingual dictionaries. Actually, a number of reviewers expressed their satisfaction with the grammatical information the OEAD provides: "not only does it give the past and present verb forms, but it also gives the plural of nouns! One can never hope to learn Arabic without this book." Unfortunately,

we noted an inconsistency in including plurals in the OEAD. While the dictionary includes the Arabic plurals of **pen**, **book**, and **nail**, the plurals of some other words such as **heart**, **needle**, and **chair** are missing.

The reason for the lack of consistency in the microstructure of bilingual dictionaries has to do with the tendency among editors to ignore the user profile. Atkins and Rundell (2008) argued that a well-defined user profile helps editors make the right decisions about both the macrostructure and microstructure of the dictionary. They put it clearly: "Know your users: that way, the dictionary will give them what they need" (2008: 28). Since this article considers the needs of the English-speaking user of Arabic dictionaries, we will evaluate dictionary treatment from this user's perspective.

The reviews for the OEAD (1972) and the COEAD (1982) refer to several other English–Arabic dictionaries. Many users suggest that Al-Mawrid and Hans Wehr (the editions are not always indicated) and even the *Arabic Practical Dictionary* by Hippocrene (2004), which is a pocket dictionary, constitute good companions as they include information which is missing in the OEAD. One Amazon reviewer states: "[The OEAD] only includes the English–Arabic dictionary, and you cannot look up Arabic text ... you also cannot look up words by roots", which suggests that some users do not typically favor alphabetic dictionaries, on the one hand, and prefer to have a bi-directional dictionary, on the other hand. We can even read reviews such as "The Hans Wehr *Arabic–English Dictionary* is probably the best you'll ever get", "If you need an English–Arabic dictionary you can't go wrong with Al-Mawrid" or "*The Arabic Practical Dictionary* by Hippocrene is much better, though it doesn't show short vowels".²

Steiner (1984: 167), who quoted Landau's comment on the dictionary reviewers, writes: "reviewers, however well-intentioned, intelligent, and in good command of the use of language, lack a basis for making informed judgments about dictionaries because they do not know why certain decisions were made [in compiling the dictionary]. They do not know what questions should be asked, much less how to answer them." Website reviews should not be totally ignored; they provide lexicographers with valuable information on the market demands. The reviewers of dictionaries on Amazon are 'actual' users who buy dictionaries and seem to know their dictionaries very well. Let us consider the number of Amazon reviews for different editions of English–Arabic/Arabic English dictionaries:

Table 1: Reviews for Hans Wehr Dictionaries

| Hans Wehr | ★ | ★ | ★ | ★ | ★ | Total |
|--|----|----|---|---|---|-------|
| <i>Dictionary of Modern Written Arabic: Arabic–English</i> (1980) | 4 | 0 | 0 | 0 | 1 | 5 |
| <i>Dictionary of Modern Written Arabic: Arabic–English Dictionary</i> (1994) | 44 | 11 | 2 | 3 | 4 | 64 |

Table 2: Reviews for Hippocrene Dictionaries

| Hippocrene | ★ ★ ★ ★ | ★ ★ ★ ☆ | ★ ★ ☆ ☆ | ★ ★ ☆ ☆ | ★ ☆ ☆ ☆ | Total |
|--|------------------|------------------|------------------|------------------|------------------|-------|
| <i>Arabic Practical Dictionary: Arabic–English/English–Arabic</i> (2004) | 17 | 7 | 2 | 3 | 3 | 32 |
| <i>Arabic Compact Dictionary: Arabic–English/English–Arabic</i> (2004) | 0 | 1 | 1 | 0 | 1 | 3 |
| <i>Hippocrene Standard Dictionary Arabic–English/English–Arabic</i> (1995) | 0 | 1 | 0 | 1 | 4 | 6 |

Table 3: Reviews for Al-Mawrid Dictionaries

| Al-Mawrid | ★ ★ ★ ★ ★ | ★ ★ ★ ☆ ☆ | ★ ★ ☆ ☆ | ★ ★ ☆ ☆ | ★ ☆ ☆ ☆ | Total |
|---|-----------------------|-----------------------|------------------|------------------|------------------|-------|
| <i>Al-Mawrid Al-Wasit: English–Arabic and Arabic–English Dictionary</i> (1985) | 2 | 0 | 0 | 0 | 0 | 2 |
| <i>A Modern English–Arabic Dictionary</i> (1997) | 1 | 0 | 0 | 0 | 0 | 1 |
| <i>Al-Mawrid Arabic–English Dictionary</i> (1999) | 11 | 3 | 1 | 0 | 2 | 17 |
| <i>Al-Mawrid English–Arabic and Arabic–English Dictionary</i> (2001) | 2 | 0 | 0 | 0 | 0 | 2 |
| <i>A Modern English–Arabic Dictionary</i> (2002) | 2 | 0 | 0 | 0 | 0 | 2 |
| <i>A Modern Arabic–English Dictionary</i> (2002) | 5 | 0 | 0 | 0 | 0 | 5 |
| <i>Al-Mawrid Al Quarib English–Arabic / Arabic–English Pocket Dictionary</i> (2006) | 1 | 1 | 1 | 0 | 1 | 4 |

Table 4: Reviews for Oxford Dictionaries

| Oxford | ★ ★ ★ ★ | ★ ★ ★ ☆ | ★ ★ ☆ ☆ | ★ ★ ☆ ☆ | ★ ☆ ☆ ☆ | Total |
|---|------------------|------------------|------------------|------------------|------------------|-------|
| <i>Oxford English–Arabic Dictionary of Current Usage</i> (1972) | 3 | 1 | 1 | 1 | 1 | 7 |
| <i>Concise Oxford English–Arabic Dictionary of Current Usage</i> (1982) | 3 | 4 | 3 | 5 | 3 | 18 |

As the tables show, some dictionaries are more reviewed than others. The Hans Wehr *Arabic–English Dictionary* (1994) received more reviews than all the other dictionaries. This compact version of the fourth edition (1980) has very positive reviews and is considered to be an essential tool for learning Arabic even by beginners. Users report that a root search in Arabic is more efficient in Hans Wehr than the classical alphabetical search. It is important to mention that the Amazon information for almost all English–Arabic dictionaries states that the customers also bought the Hans Wehr *Arabic–English Dictionary* (1994) in addi-

tion to other books on Arabic grammar. The Hippocrene *Practical Dictionary* (2004), with only 18 000 entries, ranks second. This is an indication of how much users prefer dictionaries which do not offer many Arabic equivalents for the English words, in addition to offering the IPA pronunciation of Arabic words in the A–E section. The *Al-Mawrid Arabic–English Dictionary* (1999) and the COEAD (1982) have almost the same number of reviews, though Al-Mawrid has more five star reviews.

The Al-Mawrid dictionary is probably the most cited dictionary in the literature on Arabic bilingual lexicography. The users consider it to be an authority in the field despite the fact that it lacks examples of usage. Its readability is lower due to the density in the page layout. Unfortunately we found several mistakes in Amazon.com in relation to the Al-Mawrid dictionaries. First, there is some confusion in the year of publication and the bilingual dictionary's direction (English–Arabic or Arabic–English) of some editions and added to this the fact that the exact same reviews for the *Al-Mawrid Arabic–English Dictionary* (1999) appear under another edition: *Al-Mawrid Al-Quarib English–Arabic / Arabic–English Pocket Dictionary* (2005). Such mistakes can mislead the readers who read the reviews to obtain more information on the edition they want to buy.

On the positive side, Amazon.com publishes the sales rankings of all its products, though for competitive reasons it does not publish what an item's actual sales are. The Amazon sales ranking is a good indicator of how well a dictionary is selling overall: the lower the number of the sales ranking, the higher the sales for that particular item. The bilingual dictionaries that have high rankings compared to other dictionaries in the same category are shown on the Amazon page. The table below reveals the ranking of the above-listed Arabic bilingual dictionaries. The ranking includes those dictionaries for which reviews have been written. Note that the sales rankings are updated on a regular basis and these were the results obtained on July 7, 2009.

Table 5: Sales rank of some English–Arabic dictionaries on *Amazon.com*

| | Dictionaries | Sales rankings in books (low # = higher sales) |
|---|--|---|
| 1 | Hans Wehr <i>Arabic–English Dictionary: The Dictionary of Modern Written Arabic</i> (1994) | #8 530 |
| 2 | Hippocrene <i>Arabic Practical Dictionary: Arabic–English / English–Arabic</i> (2004) | #14 441 |
| 3 | <i>Al-Mawrid Al Quarib English–Arabic / Arabic–English Dictionary</i> (2005) | #45 489 |
| 4 | Hippocrene <i>Standard Dictionary: Arabic–English / English–Arabic</i> (1995) | #89 649 |
| 5 | <i>Concise Oxford English–Arabic Dictionary of Current Usage</i> (1982) | #151 232 |
| 6 | Hans Wehr <i>Dictionary of Modern Written Arabic</i> (1980) | #298 545 |

| | | |
|----|---|------------|
| 7 | <i>Oxford English–Arabic Dictionary of Current Usage</i> (1972) | #308 359 |
| 8 | <i>Al-Mawrid English–Arabic and Arabic–English Dictionary</i> (2001) | #421 668 |
| 9 | <i>Al-Mawrid Al Quarib English–Arabic / Arabic–English Dictionary</i> (2006) | #484 053 |
| 10 | <i>Hippocrene Arabic Compact Dictionary: Arabic–English / English–Arabic</i> (2004) | #561 146 |
| 11 | <i>Al-Mawrid Arabic–English Dictionary</i> (1999) | #620 744 |
| 12 | <i>A Modern Arabic–English Dictionary</i> (2002) | #747 363 |
| 13 | <i>A Modern English–Arabic Dictionary</i> (2002) | #1 570 915 |
| 14 | <i>Al-Mawrid Al-Wasit: English–Arabic and Arabic–English Dictionary</i> (1985) | #1 763 291 |
| 15 | <i>A Modern English–Arabic Dictionary</i> (1997) | #4 425 410 |

Customers' preferences have been placed in rank order from 1 to 15. We observed the sales rankings for a few days and despite changes in the above numbers the ranking follows the exact same order. The table clearly indicates that Arabic–English and bi-directional dictionaries rank above English–Arabic dictionaries. There is also a preference for compact (less than 1 000 pages) and concise dictionaries. These results might be interpreted in different ways; it could be that a model dictionary for Arabic learners is a small bi-directional dictionary in which the Arabic words are arranged according to a root-based entry system and the pronunciation of the entry words (often in IPA) included. The English–Arabic section offers as few equivalents as possible, indicating the parts of speech of the equivalents, and presents them in concrete examples of usage. The font of the 'accented' Arabic words is readable. The reviewers' comments on the poor binding of some editions, the size and the price of the dictionary reflect the users' interest in every aspect of the book, including commercial considerations.

Notice also that the most favorably-reviewed dictionary sells very well. Surprisingly the least bought dictionary is an English–Arabic dictionary. Now whether or not we can assume a correlation between the dictionary sales rank and the number of reviews needs further investigation mainly because we have to add to our counts the number of people who read the reviews and say to what extent the reviews have been helpful. For example, 163 people have read a review for Hans Wehr (1994) alone. The review entitled, "The only essential Arabic dictionary for English speakers", has been chosen as the most favorable one. The readers said the review was very helpful. This might explain the high ranking of this particular dictionary. As for the Oxford dictionaries, 144 people have read the 25 reviews posted online: 56 readers out of 61 find the two star reviews of the Oxford dictionaries helpful while 15 out of 17 readers found the five star reviews helpful. It should be explained that the Amazon reviews are divided into 'the most helpful favorable review' and 'the most helpful critical review'. Two reviews are chosen among all the reviews to represent those two categories. For the *Al-Mawrid Arabic–English Dictionary* (1999), we can read that

45 of 45 people found the most favorable review helpful whereas 42 of 43 found the most critical review helpful.

The Amazon reviews reflect, at least to some extent, the actual situation of the English–Arabic dictionary market. However interesting, the technique presents a few challenges for the researcher. The first challenge that faces this technique is that there is no normal sampling procedure in order to do a statistical analysis. The reviewers are sometimes anonymous so we have no information about their backgrounds. Also their level of Arabic is not always self-reported. What we want to emphasize in the present article, however, is that the survey technique based on reviews on Amazon posted by anonymous dictionary users is a resource to be more fully exploited and systematized in the future. The reviewers can be asked to provide some more relevant information, a standard practice when administering questionnaires or conducting interviews. For the purposes of our analysis, we used the Amazon reviews as a point of departure in analyzing some lexical entries in the OEAD. The issue under investigation has to do with the features that can be included in a future edition of the OEAD to make it more user-friendly. As one of the reviewers put it: "Hopefully, Oxford takes the time to update and improve this dictionary."

3. Criteria for treatment of entries for English–Arabic encoding dictionaries

In an attempt to find out how helpful this dictionary is to the encoding English-speaking users, we evaluated a few lexical entries according to the following criteria:

- To what extent does the dictionary provide accurate and complete meanings of entries? Holes (1992) explained that for a lexical entry to be accurate, the listed equivalents should be distinguished and should appear in examples to help someone who is seeking *le mot juste* when translating or expressing himself in Arabic.
- Are parts of speech distinguished in Arabic? For example, the difference between the noun /sarḥ/ (edifice) and the verb /sarraḥa/ (declared) lies in the fact that a noun can be transformed into a verb by geminating the sound /r/ and by adding the vowel /a/ between the /r/ and /ḥ/ sounds. Gemination is signaled in Arabic by a symbol called 'shadda' above the sound in question [ː], and the short vowel added after the /r/ sound is indicated by a symbol called 'fatha' at the top of the sound [َ]. Derivations are very important to consider in a dictionary since derivation implies a new meaning.
- Are there usage labels which indicate the stylistic suitability of an equivalent? Many TL terms differ from each other in terms of style and register; some equivalents are rather old fashioned whereas others are in current usage, as the title of the dictionary suggests.

- Are the English cultural expressions transferred into Arabic through a literal translation, a paraphrase, or by functional equivalence? Equivalent idiomatic expressions are crucial to an English learner of Arabic who wants to use such multi-word expressions — effectively — when writing or speaking in Arabic. This is part of acquiring communicative competence in a given language. Arabic abounds with semantically challenging expressions. Does the dictionary provide the users with equivalent expressions of, at least, the most common English idioms? Which translation strategy is adopted?

Accuracy is attested in terms of sense discriminators and examples of usage, the focus being on the latter. Creamer (1987: 238) explains, "Often the information the learner is seeking, such as usage, collocations, and points of grammar, can be effectively and efficiently demonstrated by the inclusion of carefully chosen examples." It follows that the primary purpose of an example is to demonstrate the use of a word in its natural environment. Even when sense discriminators are not opted for, the example can illustrate points of usage; e.g., if the entry collocates with a certain noun indicating the typical modifiers. Examples show clearly and in an economical way the various ways the entry can be translated in context. This is exactly what an English user would expect from his bilingual dictionary: providing translations in context rather than a mere listing of the possible Arabic equivalents which as we argued earlier, suits advanced Arab learners of English who use the dictionary for decoding purposes. Consider the entry for *affect* in the OEAD in Figure 1.

Figure 1

| | |
|-------------------------------------|---|
| affect, v.t. 1. (assume) | تَظَاهَرَبِ، ادَّعَى بِ، تصنَّع، تكلف، انتحل |
| affect a beard | أَطْلَقَ لِحْيَتَهُ (متباهياً) بِهَا لَيْسَ إِلَّا |
| affect ignorance | تَظَاهَرَ بِالْجَهْلِ أَوْ عَدَمِ الْمَعْرِفَةِ، ادَّعَى عَدَمَ عِلْمِهِ (بموضوع ما) |
| 2. (produce an effect on) | أَثَّرَ فِي، حَرَكَ العَوَاطِفَ |
| it is affecting my health | هَذَا يُؤَثِّرُ فِي صِحَّتِي، هذا يضرُّ بصحتي |
| the news affected him deeply | كَانَ لِلْخَبَرِ أَثْرٌ عَمِيقٌ فِي نَفْسِهِ، هَزَّهُ النَبَأُ هَزًّا |
| 3. (concern) | هَمَّ، عَنَى |
| this doesn't affect me | هَذَا لَا يَهْمُنِي، هَذَا لَا يَغْنِينِي |

Five equivalents are provided next to each other in the first two lines without further explanation. Moreover, some words are not semantically related to the sense discriminator. Notice, for instance, the use of انتحل /ʔintaḥala/, تكلف /takallafa/, تصنع /taṣṣanaʕa/, ادعى ب /ʔiddaʕa: bi/ and تظاهر ب /taḏa:hara bi/ which, according to the editor, correspond to **affect** when it means **assume**. All the words are partial synonyms in Arabic which can be possible equivalents of **pretend** or **disguise**. The most important comment that can be made has to do with the fact that the least common meaning of **affect** is presented first instead of its most common meaning 2. Such an order can also mislead learners of English who may consult the dictionary for decoding purposes. We also question the above-mentioned words as being appropriate equivalents at all.

Let us consider another entry in the OEAD:

Figure 2

| | |
|---|--|
| add, v.t. i. (join on); <i>also add on</i> | أَضَافَ أَوْ ضَمَّ إِلَى، أَرْفَقَ، أَلْحَقَ |
| 2. (put in additionally); <i>also add in</i> | أَضَافَ، زَادَ، زَادَ عَلَى |
| 3. (put together); <i>also add together</i> | جَمَعَ، رَكَّبَ |
| 4. (calculate sum of); <i>also add up</i> | عَدَّ، جَمَعَ، حَسَبَ |
| 5. (say in addition) | أَضَافَ قَائِلًا، زَادَ عَلَى ذَلِكَ قَائِلًا |
| | additive |
| I might add | وَإِضَافَةً إِلَى ذَلِكَ، وَلِئَلِي أَزِيدُ قَائِلًا، لِي أُرْفِقَ قَائِلًا |
| <i>v.i.</i> | |
| it adds to the effect | يَقْوِي تَأثيره، يَزِيدُ مِنْ مفعوله |
| to add to my worries | عَمَّا يَزِيدُ متاعبي، وَعَمَّا زَادَ الظنَّ يَلَّةً |
| it all adds up (<i>fig.</i>) | الأدلة كلها تُشِيرُ إِلَى صحة الأمر |
| it doesn't add up (<i>i.e.</i> make sense) | هَذَا غَيْرُ معقول، هَذَا غَيْرُ منطقي، لا يُعْنِي شيئًا |
| he can't add up | إِنَّهُ لا يَعْرِفُ كيف يجمع (يُصَعِّدُهُ فِي مادَّةِ الحساب) |

The entry for **add** specifies that the verb can be transitive and intransitive. Then four equivalents are suggested: ألحق /ʔalḥaqa/, أرفق /ʔarfaqa/, ضم إلى /ḏamma ʔila:/ أو /aw/ (or) أضاف /ʔaḏa:fa/. The three last equivalents to the left do not appear later when the editor suggests the contexts in which other Arabic equivalents can be used. The verb with which **add** shares the same core meaning is أضاف /ʔaḏa:fa/, whereas the other verbs suggest additional meanings, for example, ضم /ḏamma/ which is rather **to include**, whereas the two other verbs have specific collocates which the dictionary does not mention. So no more

information is provided to help the users choose one equivalent. If we want to translate a sentence, such as, *Beat the butter and sugar together and slowly add the eggs*, only the first equivalent works out; the other words are completely irrelevant.

Under 4 (calculate sum of), three equivalents are presented without further specification: حساب/*ḥasaba*/, جمع/*jamaʿa*/ and عد/*ʿadda*/ which are near-synonyms in Arabic but are not interchangeable. For example, عد/*ʿadda*/ would be rather **to count** which does not correspond to the sense discriminator at all. It is very important to introduce these verbs in concrete examples. A native speaker of Arabic, of course, would see the nuances in meaning without any problem. An English learner, however, will certainly be confused. The same analysis applies to the entries for **adjust**, **adapt**, **affect**, **arrest** and **arrive**. Even when the dictionary provides sense discriminators in English, each sense discriminator corresponds to more than one equivalent, some of which are sometimes completely irrelevant. For example, in the entry for the noun **address**, the English-speaking user is not told how to differentiate between خطبة/*khuṭba*/ and حديث/*ḥadi:th*/, which are both polysemous words in Arabic and specific to given contexts.

Another point worth mentioning has to do with the organization of the microstructure of the entry for **add** itself. We do not understand, for example, the point behind introducing the main equivalent أضاف/*ʾaḍa:fa*/ (add) under separate meanings. The meanings of **add** which correspond to this same Arabic word should appear all together in addition to examples which illustrate the different morphological forms this verb takes: أضاف/*ʾaḍa:fa*/, ضف/*ḍif*/ (imperative) or إضافة إلى ذلك/*ʾiḍā:fatan ʾila: dha:lika*/ (in addition to that). Each form depends on the position of the word in the sentence. Such derivations, which have morphological and syntactic implications, are not really of particular interest to the editor. Unless the dictionary is intended for advanced learners of Arabic, such intricacies are difficult for most students to handle.

We also noted the use of such explanatory sentences for **ask someone out**, for example:

دعا شخصا لزيارته في بيته
daʿa: shakhsan li-ziya:rati-hi fi: bayti-hi
invite 3MS somebody to-visit-him at home-his
'to invite somebody over'

أو لقضاء السهرة معه في مكان ما
aw li-qadha:ʾi a-ssahrati maʿa-hu fi maka:nin ma:
or to-spend the-evening with-him in place some
'or to spend the evening with someone in some place'

The dictionary does not contain examples of how and where to use such equivalents which are rather explicative. Zgusta (1984: 147) writes, "The dictionary should offer not explanatory paraphrases or definitions, but real lexical units of the target language which, when inserted into the context, produce a

smooth translation." In other words, the equivalent should be a lexical unit insertable in actual sentences. Functional useable equivalents should also be semantically related to the entry word. The comparison of some entries in the *Al-Manar English–Arabic Dictionary* (1971) revealed the use of many inappropriate equivalents which would surprise a native speaker of Arabic. **Abroad**, for example, is rendered as خارج البيت /kha:rija al-bayti/ (outside the house) which is a clear mistranslation. *Al-Manar* is definitely not helpful for encoding English-speaking users as there are no sense discriminators in English for the Arabic equivalents.

The OEAD should be given credit for the fact that it *often* provides English sense discriminators for each equivalent, as it does for **add**. Each sense discriminator is put between brackets to the left of the Arabic equivalent(s): (put together), (calculate the sum of), (say in addition), etc. We noted one sense discriminator in Arabic at the end of the entry in figure 2:

لضعفه في مادة الحساب
li-ḍuʿfi-hi fi: ma:dati al-ḥisa:bi
because-weak-he at subject the-math
'because he is not good at math'

We also liked the practice of translating some expressions into Arabic such as *I might add*, *It adds to the effect*, and *He can't add up*. We thought that the last sentence could serve as a good illustrative example under (calculate sum of) instead of being presented separately. However, when asked about the grammaticality of the sentence *He can't add up*, native speakers explained that **add up** is transitive, as in *She added the bill up*, while *It doesn't add up* is an idiom, meaning *It doesn't make sense* (ironically). Lexical mistakes of this kind on both the English and the Arabic sides may now be avoided by employing corpus linguistics in dictionary creation. Arabic bilingual lexicography should develop contacts with spoken and written corpora to discover more contemporary uses of words. Lexicographers should not rely on their knowledge of Arabic because Modern Standard Arabic is nobody's mother tongue. Abu-Ssaydeh (2008) suggests the use of lexical data available through the search engine Google due to the lack of online lexical corpora. Corpus use, procedures for building up corpora and extracting data from them are all areas eagerly awaiting more research in the field of Arabic bilingual lexicography.

A general problem in the OEAD is the inclusion of some verbs as potential equivalents despite the fact that they are subject to a number of syntactic constraints. For example, when we want to translate the sentence *If you add three and four you get seven* into Arabic, we can use جمع /jamaʿa/, whereas ع /ʿadda/ would be used to translate, for instance, *He can count from one to ten*. If, on the other hand, we decide to use حسب /ḥasaba/, we need to add a noun *the sum of*, i.e., this verb cannot be used by itself. A sentence such as *Her colleagues' laughter only added to (= increased) her embarrassment*, for example, can be included to illustrate common uses of the entry word. The entry must provide an accurate

translation, or at least the relevant part of the sentence, such as, *add to someone's embarrassment* to save space. The entry for **add** includes two idiomatic expressions accurately translated, in the sense that this is how they are rendered by Arabic speakers: *to add to my worries* and *It all adds up*. Other common expressions, such as, *to add insult to injury* or *add fuel to the fire/flames* or *add salt to the conversation* do not appear under **add**. In fact, when examining other entries, we noticed that some idioms with which we are familiar are absent. The second section of this article tests the dictionary for idiomaticity. During our analysis we did not come across usage labels. The only label used in the entry for **add** is (fig.) next to *It all adds up*, which means that more usage labels are needed to provide a better understanding of the information the dictionary includes.

Apart from the fact that many equivalents are provided for each meaning, the pages in the OEAD are not cluttered and it is easy for the users to locate the meaning. Some Amazon reviewers praised the OEAD for the clarity in the page layout: "very clear and legible Arabic typeface". Each sense or example sentence is presented in a separate line. This strategy takes up space and so we suppose that the clarity in presentation is gained at the expense of the amount of information the dictionary provides. We chose some of the most common collocations and fixed expressions to test the dictionary in terms of coverage of the most typical contexts in which some high frequency words appear. We now turn to this next part of the analysis.

4. The OEAD coverage of high frequency lexical items

The above analysis suggests that scrutinizing few entries can result in pages of criticism which may "turn out to be an essay on lexicographical matters" (Steiner 1984: 167). It also revealed another problem related to the dictionary's microstructure. We had the impression that the OEAD misses frequent collocations and idiomatic expressions which are associated with some common words. It is true that no bilingual dictionary is expected to include all possible occurrences of the word and its translations. But it should at least include equivalents for the most frequent collocations, syntactic items, and fixed expressions in the users' native language to help them speak correctly or write appropriately in the foreign language. According to Shehdeh and Bin Moussa (2007: 42), "The unavailability of an expression in a dictionary very likely motivates the user to resort to either interpretation, or literal translation, which might be erroneous." Needless to say that collocations and idioms are indispensable for the encoding users who, as we saw in the Amazon reviews, are concerned with sounding natural in the foreign language.

To test the dictionary for coverage we selected a list of 113 lexical items. The list includes high frequency collocations, syntactic items, fixed expressions, and a few colloquial items. These were the lexical/grammatical areas Patzold (1994) tested in addition to some areas of vocabulary coverage in English–

German learners' dictionaries. Such a method of testing dictionaries for coverage is an attempt at a comprehensive review of dictionaries. We should like to say that our lists are not precisely similar to Patzold's as they sometimes include different lexical and grammatical items.

4.1 Collocations and fixed expressions

It can be argued that the absence of collocations from bilingual dictionaries results in the learner's use of unusual and unacceptable combinations of words. Patzold noted that the meaning 'part or piece of something' is realized differently with different nouns in English (1987: 179).

4.1.1 Noun + of + noun

(1) Cake: A slab of - (2) Coal: a bag of - (3) Cotton: a wad of - (4) Earth: a clod of - (5) Excitement: a twinge of - (6) Glasses: a pair of - (7) Guilt: a pang/ a twinge of - (8) Orange: a segment of - (9) Butter: pat of - (10) Salt: a pinch of - (11) Paper: a sheaf of - (12) Potatoes: a sack of - (13) Rubbish: a load of - (14) Snow: a drift/ flake of - (15) Tea: a pot of - (16) Water: a drink/ trickle/ sip of -

Most of the above combinations were found in the OEAD but not necessarily with the corresponding nouns. For example, the dictionary lists *a twinge of conscience* or *of excitement* instead of **guilt**. **Pang** only appears with **conscience** and **trickle** with **information**. **Flake** in (14) is used in **corn-flakes** and **soap-flakes** but not with **snow**. In (15), **tea** collocates with **bag** rather than **pot**, which is also a frequent collocation in English. Overall, we see that the dictionary generally covers such specific combinations. It also helpfully suggests other possible collocates.

4.1.2 A set of different collocational patterns

We chose some combinations from Patzold's list below (1994: 40):

(17) Amuse oneself thoroughly (18) Apologize profusely (19) Ask pointedly (20) Fall asleep (21) Mount an attack (22) Well aware (23) Wheel one's bicycle (24) Bleed freely (25) Blush furiously (26) Bored stiff (27) Cooked breakfast (28) Catch one's breath (29) The fog cleared (30) Stark contrast (31) Plodding conversation (32) Mad keen (33) Debate hotly (34) Load the dishwasher (35) Doze fitfully (36) Drink deeply (37) Call an election (38) Arouse expectations (39) Start a family (40) Heavy fine (41) Guard jealously (42) Hint darkly (43) Ignore studiously (44) Conduct an interview (45) Boil a kettle (46) Break the news

The analysis of the above expressions in the OEAD drew our attention to the problem encountered when translating certain grammatical classes into Arabic, such as adverbs. In fact, we can even assume that this part of speech is almost absent in the dictionary, as if Arabic has no adverbs. In his preface, the editor

explains that "Adverbs in '-ly' and abstract nouns in '-ness', formed from adjectives, are normally omitted unless either their formation is irregular or their usage is more frequent than or in some way different from the adjective from which they derive, e.g., 'actually'; 'muchness' [sic]" (1972: viii). The scarcity of Arabic sentences with adverbs and the absence of adverbs as possible entry words in the OEAD is a problem for encoding users.

There are a few examples where the editor of the OEAD presents equivalents for such constructions as verb + adverb, for example, **profusely** in (2) appears under **profuse** in *apologizing profusely for being late*. **Pointedly** in (3) is missing in either entry (ask and pointed), instead there is one example: *a pointed remark*. An Arabic equivalent adverb cannot possibly be derived from the adjective لاذعة /la:dhīʕa/ (a feminine adjective, as the equivalent of **remark** is feminine in Arabic). Hence, the dictionary rarely illustrates such constructions as verb + adverb as in (17), (18), (19), (24), (25), (35), (36), (41), (42) and (43). Notice that *debate hotly* is rendered as جادل /ja:dala/, a verb which includes two components as part of its meaning (debate + hotly). For example, the collocation in (20) is found in the dictionary but, as illustrated earlier in the article, three equivalents are suggested without any specification as to where to choose one or the other.

We also noted that though the dictionary does not necessarily include the above-mentioned collocations, it is a rich source of many other collocational combinations. The latter, when not presented separately, can be located in the example sentences. Two or more pages are sometimes devoted to one entry word, though some proposed equivalents appear to be uninformed by English native speakers' input which explains the Amazon reviewers' complaints reported earlier in the article.

The inclusion of more collocators would help the user choose one equivalent among the listed equivalents as in *the fog cleared* (29). The dictionary does not specify that انجلي /ʔinjla:/, which is one possible equivalent of **clear**, in fact, co-occurs with **fog**. In (38), **arouse** co-occurs with **interest** and does not appear under the entry for **expectation**. This would lead the learners of Arabic to adopt different strategies when asked to translate *arouse expectations* (38). They will probably look up **arouse** and **expectations** separately, then combine both equivalents; a strategy which does not work out all the time. We cannot look up **boil** and **kettle** in (45), then combine their equivalents since in Arabic a kettle cannot be boiled. More research on the users' strategies will reveal the way users deal with this problem and whether their strategies are successful or not.

Other inconsistencies have been identified when looking up the above expressions. One observation has to do with the presentation of labels; the latter are sometimes placed between two lists of equivalents so that we do not know where the label belongs. Of course, this would not pose a problem for native speakers of Arabic, but this would only add to the confusion of the English-speaking users. Surprisingly, **cook** in (27) is assigned an equivalent which is then conjugated in the present tense. Moreover, two alternate pronuncia-

tions, with an indication of all vowels, are suggested: يَطْهَى /yatha:/, and يَطْهُو /yathu:/.

There is also another editorial decision which seems to apply to some words and not to others. It has to do with the inclusion of some colloquial equivalents as for **bicycle** in (23) for which عَجَلَةٌ /ʕajala/ (wheel) is suggested. When looking up other entries in the OEAD, we noticed that the editor suggests some words from the dialects spoken in the Gulf or in Egypt without any indication of the dialectal variety. If colloquialisms are admitted in the dictionary there should be specific labels for them. It is our belief, however, that there is a serious need for more bilingual dictionaries in which 'current' Modern Standard Arabic is represented. A third set of phrasal verbs from Patzold will be analyzed in what follows (1987: 179).

4.1.3 Phrasal verbs + collocations:

(47) Blow up a bridge (48) Call off a strike (49) The flower has come out (50) Come up with an explanation (51) Drive out doubts (52) Enter into the spirit of something (53) Flag down a taxi (54) Gloss over a difficulty (55) Go up to university (56) Hand down an heirloom (57) Her career took off (58) Live down the past (59) Pick out a tune (60) Piece together a story (61) Set off a challenge (62) Stave off hunger (63) Talk round a subject (64) Work out a plan

The analysis of the OEAD treatment of these collocations corroborates the findings in the previous discussion. The above collocations are more specific than the combinations in the previous set due to the addition of an element (often called a particle), either a preposition or an adverb. In fact, most phrasal verbs appear in the OEAD, but not always with the above-listed collocators. The entry for **go** in (55) has three full pages which display the different meanings of the verb, in addition to a list of sense discriminators in English next to the Arabic equivalents.

The only problem is that the number of equivalents which correspond to only one sense discriminator makes the choice very difficult for the encoding user. **Come out** in (49) has five Arabic equivalents presented all on the same line separated only with a comma. In fact, each equivalent co-occurs with specific words in Arabic. To avoid repetition, the problems and the possible solutions can be summed up as follows:

- Typically, the treatment of an entry includes the use of some collocators and the omission of others, such as **suggestion** instead of **explanation** in (50).
- Given the empty space in the dictionary, we suppose that a slash between the possible, or at least the most frequent, collocators would benefit the users.
- Typically the treatment is inconsistent as to the presentation of such ex-

pressions which are sometimes presented separately in the entry as a sub-sense or as a run-on and sometimes within the translated examples.

- A dictionary, we believe, should be consistent in the presentation of information. Constant treatment of collocations in examples alone, with many more of them, would help users.

We also noted, as explained previously in the article, the use of explicative equivalents due probably to the absence of Arabic equivalents, such as **career** in (57). The absence of one-word equivalents for a given English word might logically lead to the absence of collocations, but not example sentences. Entries which explain the English words rather than provide their equivalents and possible contexts of occurrence are clearly more useful for native speakers of Arabic studying English who consult the dictionary to understand a foreign word. It seems, however, that the editor expects the users to figure out how to place such potentially confusing paraphrases in Arabic sentences on their own. Such problematic entries deserve more attention and more elaboration to suit one of the dictionary's purposes: to help active speakers and writers of Arabic.

4.1.4 Syntax items

Following Patzold (1994: 41), we selected some items from his "syntax" list to test the dictionary in terms of its provision of grammatical combinations; in other words, to test the dictionary for such subtle and important distinctions between the different prepositions which accompany a word and how these can alter meaning. These examples illustrate the way a word's syntactic behavior often provides evidence for shifts in meaning (Atkins and Rundell 2008).

(65) Correspond to sth (66) Correspond with sb (67) Die from (68) Die of (69) Furious at (70) Furious with (71) Lecture in sth (72) Lecture on sth (73) Mad about/ for sth/sb (74) Mad at sb (75) Married to sb (76) Married with (77) Relief at (78) Relief from (79) Remember + -ing (80) Remember + to inf.

Each syntactic item calls for a different equivalent, but the only available syntax items in the OEAD are (73), (74), (79) and (80). Most verbs or adjectives in English have a number of uses associated with a particular syntactic pattern. A bilingual dictionary should definitely reflect this linguistic reality. Surprisingly, **married** is nonexistent in the dictionary though Arabic has an equivalent for this English word. We suggest these combinations be presented in example sentences in case presenting them separately takes up too much space from the editor's point of view.

4.2 Other fixed expressions, metaphors, proverbs and aphorisms

Generally speaking, learners' speech in Arabic, even at an advanced level, abounds with some unusual combinations of words which are often the result

of literal translations of expressions from their native language. Fixed expressions, such as idioms and proverbs, reflect culture-specific concepts and are unpredictable from a semantic and a syntactic point of view. Their syntactic patterns constitute an integral part of their meanings. Any changes in word order or tense results in combinations which would not sound natural when heard by a native speaker of the foreign language. Arabic is rich in fixed expressions; on the other hand, many English expressions will not necessarily have equivalent expressions in Arabic.

Differences in figurative expressions between languages are due to the fact that languages do not lexicalize all concepts of the world, especially those which are not part of their social reality. It can be argued that using a bilingual dictionary is the most common strategy foreign learners opt for during the process of translating into the foreign language. Unfortunately, a translation dependent entirely on dictionary entries often results in wrong or unusual word combinations. We used a few expressions from Patzold's list of fixed expressions (1994: 39-40).

(81) Welcome somebody with open arms (82) Turn the clock back (83) Charity begins at home (84) Be a big fish in a small pond (85) Birds of a feather flock together (86) Turn a blind eye (87) Burn the candle at both ends (88) If you can't beat them join them (89) Play cat and mouse (90) Wipe the slate clean (91) A wolf in sheep's clothing (92) Curiosity killed the cat (93) Over my dead body (94) Fall on deaf ears (95) Lie down on the job (96) A drowning man will clutch at a straw (97) Come to the end of the road (98) Can't see further than the end of your nose (99) Give somebody a hard time (100) Have a short memory (101) On second thought (102) How dare you? (103) How should I know? (104) How do you do? (105) You're welcome (in response to 'thank you')

The expressions we could find in the OEAD are (85), (86), (87), (90), (91), (103), (104) and (105), i.e., only eight out of twenty five are available. It should be pointed out that the dictionary does contain other expressions (not listed above) for which it attempts to provide the most accurate equivalent expressions.

We are tempted to argue that the absence of the other expressions is due to the fact that they can be treated with paraphrases as there are no one-to-one equivalent expressions. But if the dictionary is also intended for the decoding Arabic-speaking users, as the editor claims, a paraphrase can be helpful. Arab users might even be able to provide precise equivalents on the basis of the paraphrases the dictionary offers. As for encoding users, a corpus might, as explained earlier in the article, be the only way to find more recent equivalent expressions. We also evaluated the above expressions in terms of their location in the entries. As some are listed under the first word and others are listed under words other than the first, we were not able to make generalizations. Much space in the OEAD is white space, so it is quite possible to include more of these expressions.

4.3 Usage labels

We were curious to know whether the dictionary sets out to label items according to their stylistic level. The following list, with a few modifications, was originally proposed by Patzold (1987: 174) who selected the following informal items (**bitch** has been added to the list):

(106) To smash up (107) get somebody's dander up (108) dilly-dally (109) dish out (110) a drag (111) to bum a lift (112) to dodge responsibility (113) bitch

The existence of such terms would indicate that the dictionary can help users express themselves correctly in Arabic by providing equivalents which belong to the same register. The labels for informal terms in dictionaries generally vary from 'rude' or 'slang' through 'offensive' to 'taboo'. Such labels usually indicate that the use of the term will cause offence and should normally be avoided. It is particularly useful for language learners to be warned that an item is offensive or at least informal. The OEAD indicates that (108) is colloquial and is assigned the label (coll.). **Bum** but not *bum a lift* is considered vulgar (vulg.) in the OEAD. To understand this policy better, we looked up **bitch** (noun) and there was another label attached to the word: (derog.) for derogatory in addition to a sense discriminator (immoral or malicious woman). We did not find the verb **bitch** meaning *to complain*. The other items, however, are not labeled for their informality and (107) is missing. We also noted that a word such as **crumple** is labeled colloquial, though this is not the case in any of the monolingual dictionaries we consulted. We conclude with a quote by Patzold (1991: 2967) on English–German/German–English dictionaries which perfectly applies to the current situation of the English–Arabic dictionaries: "[lexicographers] should become more aware of users' needs, especially in language production. Above all, glosses, exact and plentiful labelling, and the provision of (examples in) context give users the help they need to speak and write correct and appropriate English."

5. Conclusion

It is clear by now that decoding and encoding users of bilingual dictionaries need different kinds of help which a single dictionary cannot provide. As Holes (1992: 163) put it: "What for one user is vital contextual exemplification and explanation might be superfluous for the other user." A bilingual dictionary for English-speaking users would enable them to distinguish and pick up the most appropriate equivalents with greater confidence. As for the OEAD, English and Arabic speakers can certainly not make equally good use of the dictionary. The dictionary needs remediation at various points before it can become a useful tool for English speakers who consult the dictionary for encoding purposes. To provide more help for this category of users, the present entries need more elaboration. By elaboration, we mean more sense discriminations should be

made between the suggested equivalents. The use of paraphrases, which are rather explanatory, instead of 'real' equivalents which can fit in sentences can mislead anyone who is seeking *le mot juste* for productive use. Such explanatory phrases (potentially useful for Arab users who are decoding) should be omitted from a dictionary designed to help users in production. In order to save space, and still have a readable page layout, examples may be used to exemplify collocational, idiomatic and stylistic differences instead of presenting them in separate entries. Some 'weird' equivalents should be suppressed as they are not in common use in Arabic. The creation of a corpus would be very useful in updating the dictionary. It is also obvious from the survey based on Patzold's work that dozens of common idiomatic expressions have to be added. Encoding users need a dictionary with a more elaborate and consistent micro-structure.

We also suggest that the bilingual dictionary should include other important labels such as (written) and (spoken). There are words which are more appropriate in written Arabic and never, or rarely, used in the spoken form. A modern dictionary intended to help users express themselves correctly in the foreign language should not just be representative of words that are attested in literary usage and omit everyday usages. This practice suggests a prescriptive attitude which does not reflect how Modern Standard Arabic is actually used by its speakers in the media. A descriptive approach, reflecting the actual use of contemporary Arabic, would benefit learners. The inclusion of entries without any labels indicating usage, unfortunately numerous, can lead to awkward unnatural production. The latter can also result from the addition of regional variants, with no label for national variety, which learners of Modern Standard Arabic do not really need unless their aim is to learn a particular variety of Arabic. There are many dialectal dictionaries which describe either the dialect spoken in Egypt, the Gulf, or North Africa. Including regional variants in an English–Modern Standard Arabic dictionary is a waste of space and reflects another inconsistency in the editorial policy. We believe, however, that an extensive systematic linguistic revision of the OEAD (1972) will make it more user-friendly.

A last point we should like to emphasize again at the end of the article is that the Amazon reviews deserve more of the lexicographers' attention as they can be a decisive factor in the sales of a particular English–Arabic dictionary. Some dictionaries will be left on the shelf if poorly reviewed or not reviewed at all. It is time now that editors of English–Arabic dictionaries focus their attention on the American market. It is also important to mention that despite a remarkable increase of interest in Arabic as a foreign language in the United States, users' needs, preferences, and skills surveys are virtually non-existent. As regards the interaction of users and their dictionaries, Dolezal and McCreary (1996: 133) after having gathered hundreds of publications on language learners and dictionary use still felt the need "for more empirical, quantitative, and replicable research in the field if the results are to have a wide applicabil-

ity". When analyzing the microstructural features of English–Arabic dictionaries, their criticisms, and surveys of users' preferences, we noted that most research is directed towards Arabic-speaking users with many studies (articles and dissertations) carried out at various universities in the Arab world. Al-Ajmi (2002: 119) states: "Bilingual lexicography in the Arab world is suffering from a lack of guided practice and is in dire need of radical changes in both design and approach in order to keep pace with current advances, especially in EFL lexicography." We add to this that lexicographers need to concentrate their efforts on finding new techniques to gather information about the English-speaking users' preferences and expectations in order to meet the increasing market demands.

Endnotes

1. The 1982 COEAD claims 35 000 words and phrases. Based on the 1972 OEAD, which has only 26 726 entries, this seems unlikely.
2. For a list of the English–Arabic/Arabic–English dictionaries available in the library of Congress, see Selim (1992), who has compiled an exhaustive bibliography (213 pp.).

References

Dictionaries

- Awde, N. and K. Smith.** 2004. *Arabic Practical Dictionary: Arabic–English/English–Arabic*. New York: Hippocrene Books.
- Baalbaki, M.** 1985. *Al-Mawrid Al-Wasit: English–Arabic and Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M.** 1997. *Al-Mawrid: A Modern English–Arabic Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M and R. Baalbaki.** 1999. *Al-Mawrid Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M. and R. Baalbaki.** 2001. *Al-Mawrid English–Arabic and Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M.** 2002. *Al-Mawrid: A Modern English–Arabic Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, R.** 2002. *Al-Mawrid: A Modern Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M. and R. Baalbaki.** 2005. *Al Mawrid Al Quarib English–Arabic / Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Baalbaki, M. and R. Baalbaki.** 2006. *Al Mawrid Al Quarib English–Arabic / Arabic–English Dictionary*. Beirut: Dar el-Ilm Lil-Malayine.
- Cowan, J.M. (Ed.).** 1994. *The Dictionary of Modern Written Arabic: Arabic–English Dictionary*. Wiesbaden: Otto Harrassowitz.
- Doniach, N.S.** 1972. *Oxford English–Arabic Dictionary of Current Usage*. Oxford: Oxford University Press.
- Doniach, N.S.** 1982. *Concise Oxford English–Arabic Dictionary of Current Usage*. Oxford: Oxford University Press.

- Gaafar, M. and J. Wightwick.** 2004. *Arabic Compact Dictionary: Arabic–English/English–Arabic*. New York: Hippocrene Books.
- Karmi, H.S.** 1971. *Al-Manar English–Arabic Dictionary*. New York: St. Martin's Press.
- Wehr, H.** 1980. *A Dictionary of Modern Written Arabic*. Wiesbaden: Otto Harrasowitz.
- Wortabet, J. and H. Porter.** 1995. *Hippocrene Standard Dictionary Arabic–English/English–Arabic*. New York: Hippocrene Books.

Other references

- Abu-Ssaydeh, Abdul-Fattah.** 2008. Collocation and the Arabic–English Dictionary: Ideas for Better Dictionaries. *Zeitschrift für Arabische Linguistik/Journal of Arabic Linguistics* 48: 10-34.
- Al-Ajmi, H.** 2002. Which Microstructural Features of Bilingual Dictionaries Affect Users' Look-Up Performance? *International Journal of Lexicography* 15(2): 119-131.
- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Asfour, M.** 2003. Problems in Modern English–Arabic Lexicography. *Zeitschrift für Arabische Linguistik/Journal of Arabic Linguistics* 42: 41-52.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Creamer, T.** 1987. Beyond the Definition: Some Problems with Examples in Recent Chinese–English and English–Chinese Bilingual Dictionaries. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner*: 238-245. Tübingen: Niemeyer.
- Dolezal, F. and D.R. McCreary.** 1996. Language Learners and Dictionary Users: Commentary and an Annotated Bibliography. *Lexicographica: International Annual for Lexicography* 12: 125-165.
- Haas, M.R.** 1962. What Belongs to a Bilingual Dictionary? *International Journal of American Linguistics* 28(2): 45-50.
- Harrell, R.S.** 1962. Some Notes on Bilingual Lexicography. *International Journal of American Linguistics* 28(2): 51-61.
- Hausmann, F.J. et al. (Eds.).** 1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. Handbücher zur Sprach- und Kommunikationswissenschaft 3. Berlin/New York: Walter de Gruyter.
- Haywood, J.A.** 1991. Bilingual Lexicography with Arabic. Hausmann, F.J. et al. (Eds.). 1991: 3086-3096.
- Holes, C.** 1992. Designing English–Arabic Dictionaries. De Beaugrande, R. et al. (Eds.). *Language, Discourse, Translation in the West and Middle East*: 161-180. Amsterdam: John Benjamins.
- McCreary, D.R.** 2002. Sanseido's Concise English–Japanese Dictionary. Thirteenth Edition. *International Journal of Lexicography* 15(4): 337-342.
- Patzold, K.-M.** 1987. Context is All: The BBI Combinatory Dictionary of English and the German Learner. *Fremdsprachen Lehren und Lernen* 16: 151-182.
- Patzold, K.-M.** 1991. Bilingual Lexicography: English–German, German–English. Hausmann, F.J. et al. (Eds.). 1991: 2961-2969.
- Patzold, K.-M.** 1994. Words, Words, Words: The Latest Crop of Dictionaries for Learners of English (Part I). *Fremdsprachen Lehren und Lernen* 23: 13-64.

- Selim, G.D.** 1992. *Arabic–English and English–Arabic Dictionaries in the Library of Congress*. Washington: Library of Congress.
- Shehdeh, F. and M. Bin Moussa.** 2007. Practicality and Usefulness of English–Arabic Dictionaries in Translating English Metaphors. *Babel* 53(1): 32-47.
- Schulz, E.** 2004. *A Student Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Steiner, R.** 1984. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries. Journal of the Dictionary Society of North America* 6: 166-180.
- Welles, E.B.** 2004. Foreign Language Enrollments in United States Institutions of Higher Education, Fall 2002. *ADFL Bulletin* 35(2–3): 7-26.
- Zgusta, L.** 1984. Translational Equivalence in the Bilingual Dictionary. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 147-154. *Lexicographica. Series Maior* 1. Tübingen: Max Niemeyer.

Websites

The Amazon website (<http://www.amazon.com/>) has reviews for potential buyers of any book to read. Customers are encouraged to write reviews.

The Barnes and Noble website (<http://www.barnesandnoble.com/>) has a few reviews of English–Arabic dictionaries, usually limited to one or two per dictionary. This website also encourages customers to write reviews. Other websites for booksellers have reviews of books, but we did not find reviews of Arabic dictionaries in them.

A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography

Henning Bergenholtz, *Centre for Lexicography, Aarhus School of Business,
University of Aarhus, Aarhus, Denmark (hb@asb.dk)*

and

Rufus H. Gouws, *Department of Afrikaans and Dutch, University of
Stellenbosch, Stellenbosch, South Africa (rhg@sun.ac.za)*

Abstract: In lexicography the concepts of prescription and description have been employed for a long time without there ever being a clear definition of the terms *prescription/prescriptive* and *description/descriptive*. This article gives a brief historical account of some of the early uses of these approaches in linguistics and lexicography and argues that, although they have primarily been interpreted as linguistic terms, there is a need for a separate and clearly defined lexicographic application. Contrary to description and prescription, the concept of proscription does not have a linguistic tradition but it has primarily been introduced in the field of lexicography. Different types of prescription, description and proscription are discussed with specific reference to their potential use in dictionaries with text reception and text production as functions. Preferred approaches for the different functions are indicated. It is shown how an optimal use of a prescriptive, descriptive or proscriptive approach could be impeded by a polyfunctional dictionary. Consequently arguments are given in favour of monofunctional dictionaries.

Keywords: COGNITIVE FUNCTION, COMMUNICATION FUNCTION, DESCRIPTION, DESCRIPTIVE, ENCYCLOPAEDIC, FUNCTIONS, MONOFUNCTIONAL, POLYFUNCTIONAL, PRESCRIPTION, PRESCRIPTIVE, PROSCRIPTION, PROSCRIPTIVE, SEMANTIC, TEXT PRODUCTION, TEXT RECEPTION

Opsomming: 'n Funksionele benadering tot die keuse tussen deskriptief, preskriptief en proskriptief in die leksikografie. In die leksikografie is die begrippe preskripsie en deskripsie lank gebruik sonder dat daar 'n duidelike definisie van die terme *preskripsie/preskriptief* en *deskripsie/deskriptief* was. Hierdie artikel bied 'n kort historiese oorsig oor sommige van die vroeë gebruike van hierdie benaderings in die taalkunde en die leksikografie, en redeneer dat, alhoewel hulle primêr as taalkundige terme geïnterpreteer is, daar 'n werklike behoefte bestaan aan 'n afsonderlike en duidelik omskrewe leksikografiese toepassing. In teenstelling met deskripsie en preskripsie, het die begrip proskripsie nie 'n taalkundige tradisie nie, maar is primêr op die terrein van die leksikografie ingevoer. Verskillende tipes preskripsie, deskripsie en proskripsie word bespreek met spesifieke verwysing na die potensiële gebruik in woordeboeke met teksresepsie en teksproduksie as funksies. Voorkeurbenaderings vir die verskillende funksies

word aangedui. Daar word gewys hoe 'n optimale gebruik van 'n preskriptiewe, deskriptiewe en proskriptiewe benadering benadeel kan word deur 'n polifunksionele woordeboek. Gevolglik word argumente ten gunste van 'n monofunksionele woordeboek gegee.

Sleutelwoorde: DESKRIPSIE, DESKRIPTIEF, ENSIKLOPEDIES, FUNKSIES, KOGNITIEWE FUNKSIE, KOMMUNIKATIEWE FUNKSIE, MONOFUNKSIONEEL, POLIFUNKSIONEEL, PRESKRIPSIE, PRESKRIPTIEF, PROSKRIPSIE, PROSKRIPTIEF, SEMANTIES, TEKSPRODUKSIE, TEKSRESEPSIE

1. Introduction

The use of the terms *prescription* and *description* became popular in American linguistic studies that followed the era of Bloomfield's behaviouristic approach and American structuralism. Within this domain, these terms had a strong linguistic application. Prior to this use, the notion of prescribing, describing and even recommending had been used in scientific discussions, albeit primarily directed at textbooks. However, it is important to note that the notions of prescribing and describing did not initially result in establishing *prescription* and *description* as terms in fields outside linguistics. Although the notions of prescribing and describing were prevalent, the use of these words was of a non-terminological and general language nature. In a language like German, general language words like *beschreiben* ("describe") and *vorschreiben* ("prescribe") were rather used to carry these meanings. Unfortunately English does not have equivalents for the German *beschreiben* and *vorschreiben* other than *describing* and *prescribing* to help with the distinction between the general, i.e. the non-terminological use, and the specialised use, i.e. as terms especially in the field of linguistics. In this article, various aspects of the processes of describing and prescribing will be discussed, with the eventual main focus on the use of these words, in a non-terminological sense, in the field of lexicography. However, it will also be shown that lexicography is not the only field benefiting from an unambiguous use of these words, clearly distinguishing their use from the use of the terms *describing/description* and *prescribing/prescription* as primarily found in linguistics.

One of the real problems in the use of the contrasting pair *descriptive/prescriptive* is the fact that the distinction has primarily been in linguistics where it had a specific use and in the course of time acquired terminological status. Although the notions of describing and prescribing, used in both lexicography and some other scientific fields in a non-terminological sense, were integral components of earlier lexicographic work as well as scientific discussions regarding language criticism and textbooks, the terms *description* and *prescription* had not been integrated into the terminological collection of these fields. In the early phases of the development of lexicographic theory, prior to the emergence of lexicography as an independent discipline, many linguistic concepts, including the dichotomy *descriptive/prescriptive*, were imposed upon

dictionaries and used and interpreted from a linguistic perspective. Linguists often failed to realise that a dictionary is not an instrument that should respond to linguistic rules and criteria but is a practical tool that should be compiled in accordance with the specific needs and reference skills of a clearly identified target user.

Where linguists might have found it unproblematic to characterise their linguistic endeavours as descriptive or prescriptive, it was not so uncomplicated to describe, for instance a general language dictionary as either descriptive or prescriptive, especially not by the user of such a dictionary. Although it might have been possible to detect certain prescribing entries in a dictionary, the dictionary as such did not necessarily follow a prescriptive approach. Both description and prescription are processes. Dictionaries mostly displayed single occurrences of, for instance, prescription, because the approach had been directed at single phenomena and not at the entire dictionary. It would have been equally difficult to classify a given dictionary as descriptive, because such a classification depends on the way a lexicographer decides to present data to ensure that a function identified for the specific dictionary can be achieved. This would once again be the application of a process the user can hardly identify by merely looking at isolated instances. What one user may regard as prescriptive, another user may regard as descriptive. To illustrate this, one can look at the "grammar study note" (intended to provide information about areas of grammar tested in many examinations) attached to the article of the lemma sign *homework* in the *Longman Exams Dictionary*: "**Homework** is an uncountable noun and has no plural form. Use a singular verb after it ...". Where one user may interpret this note as descriptive because it objectively states something about the system of the language, another user may regard it as prescriptive, feeling that the lexicographer is telling the user how this word should be used.

In this article, a vital point of departure is the conviction that lexicography is an independent discipline. As a result, lexicographic practice needs to be dominated by lexicographic theory and not linguistic theory. Attention will therefore be given to ways in which the concepts of prescription and description prevail in lexicographic practice and whether they need to be introduced as fully-fledged lexicographic terms. The use of a complementing notion, i.e. proscription, and its possible terminological introduction, will also be discussed.

Before coming to the use of the concepts of prescription and description in lexicography, it is necessary to take a look at the historical development and some aspects regarding the use of these terms in both linguistics and in a more general scientific discussion.

2. Description and prescription as an old tradition

Some linguists assert that all dictionaries published prior to 1820 had been prescriptive (cf. e.g. Greimas and Courtes 1979). However, this is done without a

mutual understanding of either prescription or description. When one uses a definition comparable to that given for the German verbs *beschreiben* ("describe") and *vorschreiben* ("prescribe") the assertion is false. Statements about language, especially in grammars and dictionaries, have since their early beginnings, been primarily descriptive. Very seldom, only in a few instances, one finds items that are primarily prescribing. Nevertheless, the focus has been on these few exceptional cases, although all types of information tools usually try to reflect actual language use, as is noticeable in the following definition from Wikipedia ([http://en.wikipedia.org/wiki/Description_\(linguistics\)](http://en.wikipedia.org/wiki/Description_(linguistics))) (24.2.2010):

In the study of language, **description**, or **descriptive linguistics**, is the work of objectively analyzing and describing how language is spoken (or how it was spoken in the past) by a group of people in a speech community.

Here one should replace the word "spoken" by the word "written", because written texts, especially as reflection of the competence of the compilers of dictionaries and grammars, are authoritative. One can see in definitions like this one that the starting point is, without reservation, the observation and presentation of the observer. The purpose of this observation is not stated here. In other definitions, one can observe that such descriptions can offer help when the user of a textbook has reception problems. Others emphasise the documentary value of the description of a given synchronic section. The issue is not how someone has to speak or write in future. Yet, this remains the issue when one looks at the prevailing definitions of *prescription* ([http://en.wikipedia.org/wiki/Prescription_\(linguistics\)](http://en.wikipedia.org/wiki/Prescription_(linguistics))) (24.2.2010):

In linguistics, prescription can refer both to the codification and the enforcement of rules governing how a language ought to be used.

A reflection of the language is not at issue here; the sole issue is future language production. However, one can hold the point of view that one should speak and write in the same way as it has been done up till now or as it is most frequently done. Whether such a demand can be regarded as descriptive is questionable, because it actually perpetuates the prescription of the prevailing description. This applies, today as in the past, to different approaches and different types of needs that grammarians and lexicographers try to satisfy with their respective books. The following discussion presents a few examples taken from German grammar history, which is typical of European history. Examples of other countries and languages can be found in Haßler (2009), who gives lengthy quotations from French, Italian, Spanish and German literature of the 17th and 18th centuries. Examples from lexicographic textbooks are deliberately not used, because the purpose is to present a general view of the status of description and prescription that can then be applied to lexicography.

The first German grammars were empirical and synchronic, taking the then current German language as basis for their language description. Regarding text production, the prevailing language use was in most instances also the

example for new texts. In the earliest grammars, this was the only approach, but in the next centuries, occurred only in exceptional cases. An analysis of texts can then also lead to a criticism of certain sections of these texts. The best example comes from the most comprehensive German grammar (in five volumes with more than 3 247 pages) of Bauer (1827–1833). Bauer understood his grammar as one that directs itself at observations. These observations are actually a presentation of his own language competence, and many of his own constructed examples are used to illustrate his rules of language production. His many citations from contemporary literature merely contribute to his criticism of individual language use. Calling them "language adversities" and "sin indices", he presented them in footnotes together with lengthy comments, e.g. a four-page footnote (Bauer 1833: 333-337) with 96 citations from 58 sources, especially from Goethe and Schiller (the most famous authors of that time). Passages like these can be regarded as contributions to language criticism or simply as critical text analysis. Whether criticism can be regarded as being descriptive is a central question for the theme of this article. This question is affirmed, because it is believed that there can be no observations without judgements, including critical judgements. Bauer understood his rules as instructions for written language use: rules are deduced from the actual use. But not every form of usage has been included in the main section, with language use that can be criticised given only in the footnotes. This relation between rules and language use has since then been a basis for discussion.

A good example of arguments for and against pure description, more specifically dictatorial prescription, are known from the debate between the professors Richey and Fabricius, in papers from the year 1723. It should be noted that they did not discuss the issue as linguists. In addition, they did not provide any examples but argued quite abstractly as in philosophical discussions. Richey was an educated person and a poet. He studied Theology, Natural Science, Mathematics and History and was professor of Greek and History at a gymnasium in Hamburg. Fabricius studied Medicine and Theology. He also worked as librarian, became a doctor of Medicine and later held the chair of Theology and Ethics at the University of Kiel. Richey, whose paper has the title "Usus Tyrannus precario imperans, oder Vernunft-mäßige Sätze von der Gewalt des Gebrauches in den Sprachen / insonderheit in der Teutschen, ingleichen ob und wie weit man denselben durch Regeln der Sprach-Kunst Einhalt thun könne", argues from the assumption that no language in the world is inherent to people. From this follows that not all people learn to know a good "Mund-Ahrt" (dialect/variety) during their youth, that many do not hear the true, actual and stable use and therefore are compelled to uncertain, faulty or false language use. Richey asks whether this would not lead to a situation where the inane masses with false ways of expressing themselves gain the upper hand. Instead of following the "usus tyrannus", Richey proposes that one should carefully undertake certain changes that each reasonable person would have to employ.

It is clear that Richey wants to change or preserve the prevailing language use in order to enable reasonable discussions. Fabricius replies to these ideas in his "Usus imperium sine exceptione assertum oder die unumschränkte Gewalt des Gebrauchs / gleichwie in andern Sprachen / also auch in der Teutschen, behauptet wider die weit aussehenden Neuerungen der SprachLehrenden Malcontenten", working with the following thesis: There exists no language that has been made according to grammars or that has been improved in a uniform way. In the course of time, language has much rather been arranged by nature and all types of coincidences. Such a natural and not regulated development of a language, that Richey certainly regards as possible but also as not tenable, forms the basis for the following thesis of Fabricius and for the subsequent demands made on future linguistic work:

1. Each language is constantly subjected to changes.
2. Linguists should not restrict the deviations and changes.
3. The description of a language should rather be changed by the changed language use.

Fabricius admits that his demands on linguists imply major efforts, because the extent of language use can hardly be grasped. Fabricius regards the demand for comprehensive investigations of actual language use as the only solution. In addition the language of educated as well as uneducated people from all components of the population should be collected. When a word, construction or expression is frequently used in written and spoken language, Fabricius regards himself as not qualified to reject the form or simply condemn it. It is clear that Fabricius adheres to the famous American thesis of "leave your language alone", and therefore one should endeavour to describe real language use. It is not quite clear why he desired this. He apparently rejected every form of language criticism and as a linguist he did not want to preserve or change the structure of the language.

Fabricius writes in a fluent, elegant and especially faultless language. One can therefore assume that Fabricius would have discussed obvious mistakes. But it had not been his topic. When anyone does not write in German *Sprache* ("language") but rather *Spache* (without the *r*) every teacher, including Professor Fabricius, would regard it as a mistake that needs to be corrected. Yet, the border between **correct** and **incorrect** remains unclear. When someone writes or speaks a foreign language, the mother tongue sometimes has a definite influence that leads to mistakes. Textbooks are used to avoid such mistakes and also to be used as materials for language teaching. La Forêt (1760) is an example in this regard. This book promises in the preface to help those Danes who want to avoid the typical mistakes made when they write in German. It is not a dictionary but rather a special German grammar for Danes. The third chapter treats "corrupt words and partially bad constructions", with comments in alphabetical order. It is a monofunctional information tool, as stated in the subtitle of the work: "Lingva Germanica In Ore Danico. Das ist: Unvorgreifliche Anweisung,

Wie Ein Teutsch redender Däne Unterschiedliche Danismos in einer Teutschen Rede zu vermeiden habe: Wobey Von dem Genere der Teutschen Substantivorum gründlich und ausführlich, Von Der Declination und Conjugation zulänglich, von Der Wort-Fügung beyläuffig gehandelt wird." The brief alphabetical part of the book can be regarded as a small dictionary of false friends, as shown by the following entry:

At gaae i Caution for nogen, heisset nicht auf Teutsch, in Caution für einen gehen; sondern Bürgschaft leisten; Bürge seyn; gut für einen sagen. (**At gaae i Caution for nogen** is not in German "to be cautious for someone" but "to guarantee something, to be guarantor, to vouch for someone".)

Almost everybody would then have said, and will still say today, that "in Caution für einen gehen" does occur. And almost everybody would have said and will still say today: This is a mistake that can be explained as occurring under the influence of Danish. But when it comes to stopping and reversing a change that has already entered the language, the question is no longer whether it is *right* or *wrong* but rather whether one is trying to prevent the development of a language into something new, as was the case in the dictionary of Changuion (1844). In the preface to this restricted general language dictionary *Proeve van Kaapsch Taaleigen*, Changuion states his intent explicitly by saying: "The main purpose of the following selection ... has been to rid Dutch, in so far as the language spoken in this colony can be called that, from partially completely foreign, partially mutilated words and expressions, or at least show this direction ..." The lemmata in this dictionary represent a selection of Afrikaans words that differ from their Dutch counterparts and the treatment focuses on presenting the proper Dutch form. By doing so, Changuion consistently condemns the Afrikaans forms and prescribes the Dutch forms that he prefers and that he would want to preserve. It is important to note that his dictionary, as was the case with that of La Forêt, was not an independent publication but rather a supplement to a comprehensive grammar *De Nederduitsche Taal in Zuid-Afrika Hersteld*. The ideas Changuion discussed in the grammar were applied in the dictionary. The types of problems in the demise of Dutch in South Africa, as identified by him, were exemplified in the accompanying dictionary, which did not present an unbiased or neutral selection of items but rather those items regarded by him as contaminated. This was a clear and explicit case of language policy put to practice in lexicographic format.

A different type of language politics consists in describing specific words, expressions or syntactic constructions as wrong or ugly, and to compile a dictionary so that these wrong or ugly expressions can be substituted, e.g. Wustmann (1912). Similarly purism can be seen where foreign words are prohibited and substituted by new ones, e.g. Engel (1918).

In the 17th, 18th and early 19th century, textbooks were aids. This is more evident from the introductions of real textbooks than in the above-described

theoretical discussions. The observations can be summarised as follows from different assignments which endeavoured to find:

- (a) Description of the previous language use as a set of rules for the language system
- (b) Description of the previous language use for single phenomena
- (c) Criticism of the previous language use with real texts as examples
- (d) Suggestions for the preservation or change of the set of rules for the language system
- (e) Suggestions for the preservation or change of the use of specific single phenomena
- (f) Information as aid in text reception problems
- (g) Information as aid in text production
- (h) Information as aid in translation
- (i) Execution of language and communication policy

The last point has not been deduced from the analysis of older grammars. It has been added here to be discussed later in the article. So far, it can be determined that description had been the approach in points (a), (b), (c) and (f), whereas prescription is prevalent in (d) and (e). For language production (g), the option was either a prescriptive or a descriptive approach. Language production is therefore the point where the problem can be detected most clearly. This is not the only solution and perhaps not even the best that is found in, among others, Bauer (1827–1833) when he presents different variants but elevates one which he recommends as rule.

3. Prescription, description and proscriptio

As stated in the first sections of this article, the notions of prescription and description have prevailed in lexicographic practice, although for many years the words *prescription/description* did not acquire terminological status within the field of lexicography. To ensure clarity in future lexicographic discussions, it is necessary that *prescriptive/prescription* and *descriptive/description* as general language words but also as potential terms should be used and understood in an unambiguous way. Whereas the terms *prescriptive/prescription* and *descriptive/description* are currently used in both the field of linguistics and of lexicography, lexicographic practice has shown the need for a complementing term not previously used in linguistics but relevant in modern-day theoretical lexicography, i.e. the term *proscriptive/proscription*. Few theoretical lexicographers have given attention to the issues regarding prescription, description and proscriptio. Important guidance can be found in Bergenholtz (2003). In this section, reference will be made to some of the suggestions of Bergenholtz (2003), and these suggestions will serve as basis for further discussion.

3.1 Description

Bergenholtz (2003: 70, 71) points out the differentiated perspectives on *description* in lexicography. According to him, *descriptive lexicography* refers to the use of data from a linguistic survey and a text investigation in practical lexicographic work. He also argues that a more precise explanation is needed, because provision has to be made for different types of description. This depends on the empirical basis where the following possibilities can be distinguished (Bergenholtz 2003: 71):

- (a) introspection,
- (b) analysis of a linguistic survey,
- (c) involvement of descriptions in existing dictionaries, grammars, monographs, articles, etc.,
- (d) analysis of a number of examples randomly chosen from random texts (corresponding with the practice of dictionary making before the age of computers),
- (e) analysis of a specifically constructed text corpus, and
- (f) analysis of usage found in texts in the examined language in all available websites on the internet.

In order to deal with these possibilities, Bergenholtz distinguishes the following types of description, i.e. *open* and *hidden description*, *total* and *partial description*, *strong* and *weak description*, and *explicit* and *implicit description*.

Within the dichotomy *open* and *hidden description*, *open* refers to a situation where the outside matter of a dictionary informs the user about the empirical basis of that dictionary, whereas no such information occurs in the case where *hidden* prevails. Although many users might not be interested in it, critical users do take an interest in an open approach where the guidance in a front matter text will assist them to compare the treatment in the dictionary with their own experience of language. They need a clear indication from the lexicographer concerning the empirical basis or the investigation of the text corpus.

With regard to the distinction *total* and *partial description*, *total* implies that a combination of all the description possibilities ((a)–(f) given above) are used (cf. Bergenholtz 2003: 71). Few dictionaries employ total description. *Partial* demands that only some of the possibilities are used. Partial description can also refer to situations where the description is not directed at the dictionary article as a whole, but only at an item presenting a specific data type, e.g. an item giving pronunciation. As an example, one can refer to volume XI of the *Woordeboek van die Afrikaanse Taal* which is partially descriptive by listing all the possible pronunciation variants of a given word. As far as this one data type is concerned, this dictionary also follows an approach of explicit description, e.g. the pronunciation variants given for the lemma sign *onkonstitusioneel*:

onkonstitusioneel [ɔnkɔnstətysi^hone:l'/ɔŋkɔnstətysi^hone:l'/ɔnkɔnstətɔjone:l'/ɔŋkɔnstətɔjone:l'/ɔnkɔnstətisi^hune:l'/ɔŋkɔnstətisi^hune:l'/ɔnkɔnstətɪfune:l'/ɔŋkɔnstətɪfune:l'/ɔnkɔ:stətysi^hone:l'/ɔŋkɔ:stətysi^hone:l'/ɔnkɔ:stətɔjone:l'/ɔŋkɔ:stətɔjone:l'/ɔnkɔ:stətisi^hune:l'/ɔŋkɔ:stətisi^hune:l'/ɔnkɔ:stətɪfune:l'/ɔŋkɔ:stətɪfune:l'] b.n.w., onkonstitusionele. Ongrondwetlik: *Die konsepvoorstel vir die Wet op Surrogaatmoederskap is onkonstitusioneel en sal hersien moet word* (Beeld, 12 Febr. 1996, 6). [...]

With *strong description* an analysis of a broad empirical basis is a prerequisite, e.g. an exemplary corpus or sample, in contrast to an analysis of a narrow empirical basis in the case of *weak description*. *Explicit description* includes all results from the empirical basis in the dictionary, also, for example, obsolete words or mistakes, whereas *implicit description* includes only certain results in the dictionary and excludes, for example, obsolete words, mistakes and also informal and colloquial words and expressions. In the standard variety of English, *advertisement* is an established word. In informal English, the variant *ad* is often found. Where implicit description prevails, the form *advertisement* will be included but not the variant *ad*. Explicit description will result in the inclusion of *ad* as a lemma. Explicit description will include both the correctly spelled and the frequently incorrectly spelled forms. If the English *book* had frequently been misspelled as *boook*, a typical user of the dictionary coming across the form *boook* will need guidance to help him/her finding the correctly spelled form. Explicit description will include the form *boook*, and its treatment will refer the user to the correct form *book*.

3.2 Prescription

Bergenholtz (2003: 74) gives the following three possibilities of prescription:

- (a) a specific linguistic variant is explicitly prohibited,
- (b) one or more linguistic variants are explicitly prescribed, thus prohibiting all other non-mentioned variants, and
- (c) a specific linguistic variant is explicitly prescribed (as opposed to prescription (b), this involves a new word, new spelling, new pronunciation, new inflection or neologism).

It should be noted that prescription is typically employed in specific dictionary articles, and therefore all these possibilities are not used in every article of a dictionary with a prescriptive approach. The distinction between a descriptive and a prescriptive approach is not always absolutely clear, and dictionaries

opting for a prescriptive approach will employ this approach in certain articles but will also include articles similar to, for example, those in weakly descriptive dictionaries. As is the case with description, prescription can also be explained with reference to different dichotomies, i.e. *open* and *hidden prescription*, *total* and *partial prescription*, *strong* and *weak prescription*, *explicit* and *implicit prescription*. With *open prescription* the outside matter contains a note in which the intention of the dictionary to influence and eventually change the language use is indicated, whereas such a note does not occur in the case of *hidden prescription*. One of the best known "declarations of intent" in this regard is Samuel Johnson's passion for "fixing the language", as expressed in *The Plan of a Dictionary of the English Language* (1747) that preceded the publication of *A Dictionary of the English Language* (1755). It has to be added that Johnson later admitted that he eventually had to abandon the idea of his dictionary succeeding in fixing the language.

In *total prescription* a combination of the prescription possibilities (a)–(c) is used, while in *partial prescription* only one or two of the possibilities can be used. Partial prescription can also refer to situations where the prescription is not directed at the dictionary article as a whole, but only at an item presenting a specific data type, e.g. an item giving pronunciation or an item giving the orthographic form of the lexical item. The original meaning of the Afrikaans word *sondebok* ("scapegoat") was "person unfairly blamed for wrongdoings". In everyday language use, this word acquired a second sense, i.e. "person who trespasses". From a puristic point of view, linguists and lexicographers warned against this new sense, and a partial prescriptive approach could have seen a treatment directed at indicating the negative attitude towards this second sense.

Strong prescription prohibits and only allows certain variants in cases with a clear difference from the normal language, whereas with *weak prescription* the dictionary articles have items that prohibit and only allow certain variants in certain cases without showing such a clear difference from normal language use. With *explicit prescription* one variant is explicitly allowed and another explicitly prohibited, whereas *implicit prescription* allows one variant but prohibits all other variants by omitting them.

3.3 Proscription

Proscription allows the same possibilities for the empirical basis as description, cf. the options (a)–(f) given above in section 3.1. However, the results of empirical analysis are dealt with in a different way compared to a descriptive approach. In this regard the most salient distinction lies in the fact that the lexicographer does not only provide the results from the empirical analysis but goes further by indicating a specific variant that he/she regards as the recommended form. Bergenholtz (2003: 77) sees it as "selective description", a formulation which has to be interpreted as referring to weighed description. The

following distinctions apply to proscription: *open* and *hidden proscription*, *total* and *partial proscription*, *strong* and *weak proscription*, and *exact* and *non-exact proscription*.

Within the dichotomy *open* and *hidden proscription*, *open* refers to a situation where the outside matter of a dictionary informs the user about the empirical basis of that dictionary, whereas no such information occurs in the case where *hidden* prevails. With regard to the distinction *total* and *partial proscription*, *total* implies that a combination of all the proscription possibilities ((a)–(f) given above) are used, whereas *partial* demands that only some of the possibilities are used. With *strong proscription* an analysis of a broad empirical basis is a prerequisite, in contrast to the analysis of a narrow empirical basis or, in the case of *weak proscription*, without any reference to a broad empirical basis. *Exact proscription* demands that only one variant is recommended although other variants may be mentioned, whereas *non-exact proscription* recommends more than one variant, while other variants may be mentioned. Volume XIII of the *Woordeboek van die Afrikaanse Taal* shows a change in approach with regard to the presentation of items giving the pronunciation of a word. In the previous volumes, a descriptive approach resulted in all the variants being given without any recommendation from the lexicographer. Volume XIII still offers all the variants, but the user guidelines state quite emphatically that where more than one pronunciation variant is given, the first is the variant recommended by the lexicographer. This is a type of exact proscription. Complementing the notion of exact proscription where other variants may be mentioned (Bergenholtz 2003: 77), Gouws and Potgieter (2010) make provision for an approach of *exclusive proscription* according to which only the recommended form and not the other variants will be included in a given dictionary. However, unless a form of marking or labelling is used in the dictionary, the user will be unable to distinguish between exclusive proscription and implicit prescription. Important in a proscriptive approach is an acceptable way to deal with the non-recommended forms. Tarp and Gouws (2008) discuss various aspects regarding the implementation of a proscriptive approach, including different ways in which the non-recommended forms can be presented.

In the existing literature on proscription, the defining feature has been identified as the giving of a recommendation. This remains the unique characteristic nature of proscription. However, the application of a proscriptive approach can be enhanced if the recommendation is supported by a motivation in which the lexicographer gives his/her reasons for opting for the specific variant or form and not for another. Such a motivation may give the user the needed assurance regarding the appropriateness of the specific form.

3.4 Realising ideological issues

The use of prescription, description and proscription in dictionaries should not be seen in isolation. As the motivation for these approaches often lies on a dic-

tionary-external level, these applications can be part of the attempted realization of an underlying ideological issue. Two of the real issues coming to the fore in the application of, for example, a prescriptive, descriptive or proscriptive approach are language policy and language criticism. Bergenholtz and Gouws (2006) have argued that every lexicographic decision may be regarded as the result of a political decision, more precisely a language and communication policy decision. Dictionaries can be employed to support a specific language and communication policy, and a prescriptive, descriptive or proscriptive approach may be chosen to realise such an assignment.

3.5 Description, prescription, proscription: Opting for the best practice approach in general language dictionaries

Although Bergenholtz (2003) has indicated that in total description/prescription/proscription, a combination of description/prescription/proscription possibilities are used, he did not make provision for a combination of subtypes of description/prescription/proscription. However, before discussing this possibility, it is important to know exactly when prescription/description/proscription could be used in dictionaries. Yet again the determining role of lexicographic functions needs to be acknowledged. Within a user-driven approach to lexicography, every aspect of the contents and structures of dictionaries have to be determined by the envisaged function(s) of the specific dictionary. The central question the lexicographer needs to ask him-/herself remains: "What do I want my user to be able to do with this dictionary?" A functional approach demands that notions like prescription, description and proscription should be seen as ways to assist in achieving the function identified for a given dictionary. Consequently the decision of a lexicographer should not be to compile, for example, a prescriptive, a descriptive or a proscriptive dictionary, but rather to compile a dictionary to achieve for example a text reception, text production or cognitive function. This function of the dictionary will then determine a process that might lead to a stronger prescriptive, descriptive or proscriptive approach.

Having identified the function of a dictionary, the lexicographer may decide on a process with a stronger prescriptive, descriptive or proscriptive approach. The lexicographer will be aware of the specific process employed in the given dictionary, but, although other lexicographers may endeavour to detect the approach, they will find it difficult to make an unambiguous classification, although they may be able to identify certain cases of prescription, description or proscription in the treatment presented in some individual articles. Where text production is the prevailing function, the lexicographer may do well to introduce some prescriptive or proscriptive entries in order, for example, to ensure innovative language use, to preserve existing variants or to bring about changes. Where text reception is the envisaged function, a prescriptive approach serves no purpose. The lexicographer could assist the user

much better by introducing a descriptive approach based on a thorough text analysis.

When opting to compile a dictionary based on lexicographic functions, the lexicographer is not primarily interested in the different subtypes of prescription, description or proscription, valid as they may be. He/she much rather tries to employ a combination of these subtypes that can best help to achieve the identified lexicographic function. This demands the application of open, total, strong, explicit description, prescription or proscription. Such a combination helps the lexicographer better to achieve the desired function.

4. Reception

The principle that all lexicographic decisions should be determined by the envisaged lexicographic functions, also applies to the methods of selection, i.e. whether a descriptive, prescriptive or proscriptive approach is followed. Selection according to functions at the same time assumes that a quicker and better access of the data in the dictionary can be obtained in an unproblematic way.

There is, for example, no access possible when an orthographic variant does not belong to the dictionary data. Until 2001, two Danish equivalents for the English word *line* were used and officially recognised by the language commission, i.e. *linie* (with *-i-*) and *linje* (with *-j-*). But since November 2001, only *linje* (with *-j-*) has been permitted. In new dictionaries compiled after this date, and this applies to both printed and internet dictionaries, only the permitted variant has been included. The newly prohibited form does not form part of the dictionary data. Because this variant previously was and still remains the most frequently used form, many users are looking for guidance regarding this spelling. This type of prescription therefore has negative consequences for the usability of a dictionary. A prescriptive approach that omits the prohibited variants, does not give adequate solutions when users are looking for an aid that can really assist in solving reception problems. Proscriptive solutions are also unsatisfactory and not recommended, because information regarding the meaning of words, word combinations or sentences is needed when reception problems are experienced.

For a general language reception dictionary, an open, total, strong and explicit description would usually be the optimal procedure. For specialised reception dictionaries, a text corpus is not required as empirical basis, because only partial and not total description is needed. After it has been established that reception dictionaries need to be descriptive, the specific consequences still have to be resolved. The types of items required in a reception dictionary that have to be selected and described according to descriptive methods still have to be established. Therefore the following suggestions are given albeit that the last two ((c) and (d)) occur in general language dictionaries, but very seldom in specialised dictionaries:

- (a) Lemma
- (b) Item(s) giving the meaning of the lemma
- (c) Idioms and other word combinations whose meaning cannot be determined by merely looking at the meaning of the individual words
- (d) Item(s) giving the meaning of idioms or other word combinations

Regarding the selection of lemmata and idioms, it should especially be noted that the frequency criterion, so popular in modern-day lexicography, does not have priority. In prevailing log file analysis (cf. Bergenholtz and Johnsen 2007), no investigations have been made to determine whether the words with the highest usage frequency in texts are also those consulted the most frequently for text reception problems in internet dictionaries. It is presumed that the words most frequently looked up in general language dictionaries for reception problems belong to the following groups: (a) infrequently used words, (b) obsolete words, and (c) polysemous words (where the user is not quite sure how to distinguish between the different senses). But frequently used words are certainly not looked up so often, at least not by mother-tongue speakers. Amongst foreign users, it would be different. In specialised dictionaries, there would be big differences between lay persons, semi-experts and experts. When consulting a reception dictionary, lay users would typically behave as foreign users, i.e. they would often look up the frequently used terms. However, although these are still only suppositions, they seem to apply more likely when selecting lemmata and idioms, than the highly acclaimed frequency criterion (cf. Bergenholtz 1992, 1994).

There is an additional selection possibility resulting from a late survey of test persons. A survey is regarded as one of the empirical possibilities to gather data for description. This possibility is familiar from the experience with printed dictionaries where interested users write to the publisher that they looked up a word or expression but could not find it in the dictionary, constituting a so-called lemma gap. In electronic dictionaries, this method can be optimally employed because the log files can be examined every day for the inclusion of those words and expressions, looked up but not found, as lemmata in new dictionary articles.

Those words and idioms presented in reception dictionaries have to be provided with an explanation. This is the item that the data users look-up must contain to ensure a successful consultation. But how does a descriptive item giving the meaning look? In the first instance, all senses of a lemma or idiom should be given and no "non-fitting" senses should be omitted, e.g. the second sense of the Danish verb *hustle*, meaning both "to make an untrue and deceitful statement" and "to have sex with a man for money or other favours". This might perhaps not be controversial. But a descriptive item presenting meaning may also consist of critical remarks regarding specialised terms, e.g. from a scientific perspective it is false for a music dictionary to explain the term *Vienna Classicism* as referring to Mozart and Haydn, because they belong to the classi-

cal period. This then is not a prescriptive item but a describing addition that aims to correct the scientific classification.

It can certainly be argued that collocations, examples or synonyms can also be helpful when a reception problem occurs. It has to be stressed that the user finds the collocations and examples in the text where the reception problem occurs. The user therefore needs an appropriate item giving the meaning that will solve the reception problem. It here concerns the principle of a monofunctional dictionary containing as much data as necessary but as little as possible to guarantee a rapid access that is not impeded by unnecessary data or that leads to information stress or even information death.

Finally, by the use of the predicate *descriptive*, it has not been indicated how comprehensive the item giving the meaning in a reception dictionary should or could be. The question really is whether a descriptive item giving the meaning should only contain so-called essential items in order to be an optimal entry in a reception dictionary. This does not imply a direct distinction between semantic and encyclopaedic data — a distinction that in any case cannot be maintained from a purely scientific perspective (Haiman 1980: 351 and Bergenholtz and Kaufmann 1996). The question of what is needed and of what is redundant is important, because a pure descriptive description of the meaning of a word can easily amount to the extent of a complete book. The distinction between essential and non-essential items giving meaning was a central theme in the journal *Dictionaries* (1993).

Wierzbicka (1993) distinguishes and believes to be able to distinguish clearly between meaning and knowledge. She argues that much space is wasted when encyclopaedic knowledge is also presented in a dictionary article. She criticises the following article:

sugar a sweet substance that consists wholly of sucrose, is colourless or white when pure, tending to brown when less refined, is usually obtained commercially from sugar-cane or sugar-beet, and is nutritionally important as a source of carbohydrate as a sweetener and preservative of other foods.

and proposes instead a lexicographic definition that is much easier to understand:

sugar something that people add to things they drink or eat when they want to make them taste sweet; it comes from some things growing out of the ground; it is white.

To both articles various entries could be added, e.g. that Wierzbicka's proposal does not contain "sugar-cane" or "sugar-beet" but rather (a) "it comes from some things growing out of the ground" (b) "it is white".

The latter statement (b) is false, because brown sugar is also found. The former part (a) is difficult to comprehend. Even a young school child would understand "plants" and "sugar-cane" rather than the formulation the author obtained by means of a linguistic theory of semantic primitives and not a the-

ory of dictionary functions. The real question in this regard is that of description. Both these entries can hardly be called descriptive as far as descriptive completeness (cf. Bergenholtz and Schaefer 1977) is concerned. Comprehensive books have been written about sugar. The real issue is not whether it is comprehensively descriptive or in one way or another prescriptive, but rather for whom and for what purpose the meaning has been presented in the dictionary. It can be maintained that for receptive needs a descriptive approach is the only appropriate one. But there are different ways of presenting an appropriate description.

Whether long or short is not the question, but rather for which user group and which type of usage situation the entry has been planned. Also the distinction between semantic and encyclopaedic is not really the issue, because this distinction is scientifically not tenable (cf. Bergenholtz and Kaufmann 1996). The question is rather whether assistance is needed for reception, for obtaining as much knowledge about a given matter, word or term, and also whether for lay persons or semi-experts. Compare in this regard two different articles, both planned for a reception dictionary. The first is aimed at lay persons in the specific field:

bacteriophage Bacteriophages belong to a group of viruses that infect bacteria.

The second entry comes from a proposal for a reception dictionary for semi-experts in the given field:

bacteriophage Bacteriophages, or phages, are viruses that infect bacterial cells. The size of a phage particle is 20-200 nm (1 nm = 10^{-9} m). A phage usually consists of two components, a chromosome of DNA or RNA and a protein coat, the capsid, which serves as a protective shell containing the genome and which is involved in the infection process.

The extent of the articles is not the question, but rather which items are needed for a reception of different user groups of a given dictionary. Bergenholtz (1998) also contains the following test entry for a gene technology dictionary for lay people and semi-experts. Here the entry for semi-experts is quite brief, because this term is not often used, compared with the terms *gametes*, *gamete-producing cells* and *bacterial spores*:

germ cell The expression germ cells is sometimes used for gametes, gamete-producing cells, or bacterial spores.

The test entry for lay people is in principle not only longer, but also more informative, because an item giving the meaning is presented corresponding to the general language use, i.e. how *gamete* is used:

germ cell Germ cells are the base for sexual reproduction; a fusion of a male and a female germ cell causes the fertilization of the egg.

5. Text production

5.1 General remarks regarding text production

In the previous section on text reception, it has been mentioned that this article is directed at monofunctional dictionaries. This needs to be emphasised yet again, because of the real implications a monofunctional approach also has for text production. Too many existing general dictionaries give no indication regarding their specific communication function. The default approach of many lexicographers is that they are producing a polyfunctional dictionary which should assist the user in satisfying at least a cognitive function, a text reception function and a text production function. Consequently dictionary users do not really know what a general language text reception or text production dictionary should look like, because they have not encountered such dictionaries in their usual dictionary consultation procedures. Where such monofunctional dictionaries have been produced, users responded in a positive way and utilised these dictionaries as practical instruments to assist them in solving specific problems. When text production is focused on from a descriptive/prescriptive/proscriptive approach, it is important to have an unambiguous interpretation of the implications of text production for the content and presentation of dictionary articles. One of these implications is that the dictionary will not necessarily include all the high-frequency lexical items from a given language nor will the dictionary necessarily display a homogeneous article structure or a comprehensive treatment of the meaning of a given word. A detailed discussion of the choice of lemma candidates and the data types to be included will not be given in this article — it belongs in an article on text production. However, some aspects of these issues should be dealt with here.

With regard to almost all aspects of text production, a distinction needs to be made between dictionaries compiled for mother-tongue speakers of the treated language and dictionaries compiled for non-mother-tongue speakers of the treated language. The scope of the data included in a text production dictionary will also be determined by the medium of the dictionary, i.e. whether it is planned as a printed or an electronic dictionary. This aspect will not be discussed in this article.

The following types of items could be included in a general language text production dictionary: the lemma sign, pronunciation data, grammatical data, a brief explanation of the meaning of the lemma or a mere translation equivalent, one or more example sentences, one or more collocations, synonyms, and antonyms. Where idioms and fixed expressions are included as treatment units, a text production function compels the lexicographer to present items indicating the typical use of the idioms and fixed expressions.

It is important to pay attention not only to the data types that have to be included in, but also to those items or indicators that should be omitted from a text production dictionary. A popular feature especially in English learners'

dictionaries is the use of non-typographical structural indicators to mark usage frequency of words, e.g. a system where five stars indicate that a word belongs to, say, the top 500 words of the language, whereas one star indicates that it belongs to, say, the top 3 000 words of the language. From a text production perspective, these indicators have little value. The user consults the dictionary to find text production assistance regarding a specific word. When he/she finds this word in the dictionary, the information that needs to be retrieved is related to the use of the word. A mere indication of the usage frequency data, i.e. usage frequency for the sake of usage frequency, does not help with text production. The only value it may have could be if it is used in a relative way, i.e. to support other entries. For example, in the articles of words marked as infrequently used, a cross-reference is given to an appropriate synonym with a higher usage frequency that could enhance the communicative success of the text production procedure. Another value of frequency indications could be to support the motivation for a given recommendation when applying a proscriptive approach. For example: Danish has the following sets of variants: *bevislig* x *beviselig* and *ubevislig* x *ubeviselig*. The proscriptive Danish internet text reception dictionary (*The Danish Internet Dictionary*) recommends the use of *bevislig* and *ubeviselig*. Following an enquiry from a user about the choice of recommended forms, an investigation of the usage frequency of these forms indicated that the recommended forms show 9 180 and 2 208 occurrences respectively, whereas their non-recommended counterparts show a usage frequency of 1 900 and 448 respectively. This support from a corpus enabled the lexicographer to add a motivation for the recommendation in which usage frequency is given as a criterion for the specific recommendations.

In a text production dictionary, the treatment of the lemma has to be planned in terms of the needs of the user of the specific text production dictionary. The criteria for the selection of lemma candidates for a text production dictionary have to differ from the criteria applicable to text reception dictionaries. The typical needs of mother-tongue speakers regarding the selection of lemmata will differ from those of the non-mother-tongue speakers. Lexicographers can work with the assumption that mother-tongue speakers will have a better knowledge of high usage frequency words compared to non-mother-tongue speakers. However, where this will play a determining role in the selection of lemma candidates for a dictionary for text reception, it will play a diminished role in text production dictionaries. The user of a text production dictionary does not primarily need to obtain the meaning of the common word but rather information regarding its productive use. High usage frequency words may have unpredictable uses or collocations for a given polysemous sense that are not so well known. In order to present the user with these example sentences or collocations, even high usage frequency words may qualify as lemmata in text production dictionaries. The selection of lemmata must be determined by the text production assistance the inclusion and treatment of a given lexical item as lemma can give to the target user of the specific diction-

ary. This implies that not only correctly spelled words should receive lemma status. Words that are frequently misspelled should be considered for inclusion as lemma candidates. A user might want to use a word in a text production situation, but is uncertain of the spelling. Looking for the incorrectly spelled form (cf. the comments on the form *boook* in section 3.1) and finding it in a dictionary with a cross-reference to the correctly spelled form, the user will be assisted in proper text production. The following can serve as another example: The German equivalent for *sixteen* is *sechzehn*. Users often think the spelling is *sechszehn*. This incorrectly spelled form can be included, with a cross-reference to the lemma sign *sechzehn*. Having found this lemma and the cross-reference in the dictionary, the user can use the correct spelling *sechzehn* when producing new texts.

For text production purposes, pragmatic guidance can be very helpful. A text production dictionary should therefore employ a well-defined set of pragmatic labels to guide the user with regard to, among others, stylistic, geographic and subject-specific restrictions of a given word, expression, sense or other item in a dictionary article. The use of stylistic labels like *informal*, *colloquial*, *obscene* and *vulgar* constitutes a type of guidance non-mother-tongue speakers of the treated language especially need to avoid communicative embarrassment.

Data on pronunciation is valuable for spoken text production. Where a dictionary has oral language use in its text production scope, items giving pronunciation should be included. This is primarily as a response to the needs of users who are non-mother-tongue speakers of the language treated in the dictionary. For mother-tongue users, more limited pronunciation guidance could be given. The treatment in text production dictionaries is often focused on written texts, and in these dictionaries pronunciation guidance is not needed. Where the focus is on written texts, grammatical data, e.g. entries presenting pluralisation, degrees of comparison, and the tenses of verbs, can play an important role. The user might be familiar with, for example, the singular form of a noun, but needs to use the plural form in a given text. In a dictionary for mother-tongue speakers, items giving the plurals that are formed systematically are not so important, because the typical user should be familiar with the systematic and predictable morphological patterns of the language. Again, however, it is important that the lexicographer has to work with a clearly identified target user and has to be familiar with the needs and reference skills of this user. The absence of items giving the plural forms of some nouns should not imply that no plural form may be given. Where the lexicographer regards it as important for the target users in their text production endeavours to have access to a specific, typically unsystematically formed plural, such an item needs to be included. It should be noted that, for example, all nouns do not have to be treated in exactly the same way in such a dictionary. The lexicographer should employ the users' guidelines text to give an account of the different approaches in the dictionary. However, the lexicographer should never use

the opportunity to give explanations in a users' guidelines text as an excuse to present the data in the dictionary in an ambiguous way or in such a complicated or condensed way that the user can only comprehend the system of the dictionary by using the guidelines text. A good dictionary will include a users' guidelines text, but the presentation in the articles has to be of such a nature that the typical target user should be able to use the dictionary, albeit not necessarily in an optimal way, without consulting the users' guidelines text.

Where users have text reception problems, the typical information they need to retrieve from a dictionary article is of a semantic nature. They need a relatively comprehensive explanation of the meaning of the word. Contrary to this, the users of a text production dictionary merely need to confirm the meaning or a specific polysemous sense of a word. A full explanation of the meaning is not needed. If the dictionary is compiled for non-mother-tongue speakers, the semantic data may be restricted to a translation equivalent in the mother-tongue of the target user. This will confirm the meaning of the word to the users and enable them to proceed with their text production activities. In addition to the explanation of meaning and the presentation of translation equivalents, an indication of semantic relations, especially synonymy and antonymy, holding between lexical items is a typical part of the semantic assignment of dictionaries. The inclusion of synonyms in a text production dictionary assists users in finding a more varied selection of words. Because the user is consulting the dictionary by looking at the treatment of a specific lemma, the synonyms should be included as lemmata in their own alphabetical article stretches and each article needs a cross-reference to the other synonym(s). The explanation of meaning will typically be given in the article of one of the synonyms, with the other article(s) containing a cross-reference to this article. In the distinction between absolute and partial synonyms, it becomes clear that partial synonyms occur much more frequently than absolute synonyms. The proper use of partial synonyms is more demanding than that of absolute synonyms. Compared to non-mother-tongue speakers, mother-tongue speakers are better equipped to use them appropriately. Consequently a text production dictionary for mother-tongue speakers of the treated language may embark on a more comprehensive presentation of partial synonyms than its non-mother-tongue speaker counterpart. In text production, the negation of a contrasting form is often employed to express a given meaning. Instead of saying *the door is open* one would say *the door is not closed*. In order to enable users to express themselves in this way, it is important that text production dictionaries should also include some antonyms.

One of the most salient features of a text production dictionary is its presentation of example sentences and collocations. It has to be accepted that only a limited number of examples and collocations can be included in a printed dictionary. An electronic dictionary has less space restrictions, but can also not include everything. In dictionaries claiming to have a text production function, it is often found that the examples illustrate the most general and typical use of

a given word, and the collocations are the most frequently encountered ones. Again the lexicographer should negotiate the target users. Where non-mother-tongue speakers of the treated language are the target users, some of the general collocations and example sentences giving typical uses can be included, but the mother-tongue speakers are usually familiar with the most typical uses and collocations and are rather in need of lesser known and used forms.

In trying to satisfy a text production function, the lexicographer should pay careful attention to the application of an approach characterised by either description, prescription or proscriptive or a hybrid application in which more than one of these approaches can be combined. This decision should not be made in a haphazard way.

5.2 Text production: Description, prescription, proscriptive?

Where prescriptive and proscriptive approaches are employed, a user consulting a dictionary for text production purposes should preferably know whether the lexicographer is prescribing or recommending a given form and by doing so signalling that there are other non-prescribed and non-recommended variants. Where a given form has no variants, e.g. regarding pluralisation or orthography, the form given in the dictionary article is not only the only appropriate form but often also the only existing form. Users need to be informed accordingly so that they can know the entry does not represent an application of implicit description, implicit prescription or exclusive proscriptive where only one variant is included and the others excluded. But even when only one recognised form exists, one often finds non-recognised forms that are the result of, for example, spelling mistakes. As indicated earlier in this article, a text production dictionary should make provision for the inclusion of these forms to assist the user in not using them.

When one looks at the different types of description, prescription and proscriptive (cf. Bergenholtz 2003 and section 3 of this article), it becomes clear that *description* is not a viable option for text production if more than one variant prevails. The user in need of text production assistance does not want to make choices, but is rather looking for the best form for a given context. This applies to the choice of a word or a grammatical form, e.g. the choice between two plural forms, synonyms, collocations, etc. This is especially true in the case of users who are non-mother-tongue speakers of the treated language and do not have the ability to evaluate the different variants and their appropriateness for a given situation of use. The success of their dictionary consultation procedure should not rely on an uninformed choice they make between different variants. Mother-tongue speakers of the treated language who have a good command of their language will also have to make a choice when confronted by a descriptive approach in which more than one variant is presented. Even then description will not be the ideal solution, especially not for a quick answer to their text production problem, because negotiating the different variants can

be time-consuming. However, where cross-references between lemmata representing different variant forms are given, a knowledgeable user may benefit from the descriptive approach by being made aware of different variants from which a choice could be made. If these variants are not all equally suited for all contexts, they have to be labelled to indicate, for example, style and usage differences. Although the presentation of variants can be seen as a form of description, the added labels bring a proscriptive or even prescriptive nuance to the fore, because they indicate the environment where the use of the specific variant is prescribed or recommended. This can be regarded as another form of hybridisation between description, prescription and proscription.

For text production purposes, *prescription* can be a viable option. A prescriptive approach influences future text production activities and gives the user one prescribed form to use. The success of prescription in a text production dictionary depends on the type of prescription employed. Implicit prescription which allows one variant but prohibits all the other variants by omitting them is of little assistance, because the users will not find the lemma they are looking for if it is a prohibited variant. Explicit prescription with one variant explicitly allowed and another explicitly prohibited might be more useful. A cross-reference entry should guide the user from the prohibited to the prescribed form. Prescriptive entries are often guided by ideological and language political motivations. They do not always represent the actual and default language use. Following the prescribed advice can lead to a user producing a text in which unnatural language is used, e.g. superstandard, puristic or idiolectal forms. Where a user relies on prescriptive dictionary entries, it may impede the communicative success of a text production process. However, if the lexicographer wants to introduce an innovative form and assist users in using such a form in a proper way, the prescriptive approach has its advantages.

Because *proscription* relies on an empirical basis, the actual language will be reflected in a proscriptive approach, especially in an application of strong proscription where an analysis of a broad empirical basis is a prerequisite. Explicit proscription gives the lexicographer the opportunity to include different variants, but also to indicate the recommended form. If a proper system of cross-referencing is employed to link the non-recommended variants to the recommended form, proscription can be the best option in a text production dictionary. This applies to all the different data types where variation occurs and a recommendation is made.

A proscriptive approach can add value to the items presented in the article slot for collocations. A given word can combine with different words in different collocations in order to express the same meaning. In Afrikaans the noun *antwoord* ("answer") can be used in the collocations *'n antwoord gee*/*'n antwoord verstrek* ("give an answer"). As both are correct and good Afrikaans, a descriptive approach will include them as variants. The collocation *'n antwoord verstrek* is slightly more formal than *'n antwoord gee*, but not so formal that it needs to be

labelled as such. The latter form can be used in both formal and informal situations and should be recommended as the best option for general use. Here the application of exact proscription will not only allow the inclusion of the recommended form, but also that of *'n antwoord verstrek* as a variant form.

A major advantage of a proscriptive approach lies in the fact that the lexicographer as someone acutely aware of the needs of the intended target users of the specific dictionary, can make a recommendation that should suit the intended target user in the best possible way in his/her text production endeavours. In this regard it is important to note that a proscriptive approach does not imply that the same recommendation for the same function will necessarily be given in different dictionaries. The recommendation in a text production dictionary for learners in primary school may differ from the recommendation in a text production dictionary for adult users. The needs of the user will determine the nature and extent of the recommendation to ensure the best text production possibility.

6. Mono- and polyfunctional dictionaries

This article is not exhaustive, because only two of the four main types of dictionaries with communication functions have been discussed. The remaining functions of text correction and translation demand a separate discussion. Even more important would be the discussion of the cognitive functions, which have not been treated here. Such a contribution is urgently needed because of all the uncertainty and vagueness that prevail. However, some aspects in this article do apply to the cognitive functions of a descriptive dictionary (e.g. a documentary function), some to that of a prescriptive dictionary (e.g. when aiming to achieve a fundamental or partial change of existing norms or uses) and some to that of a proscriptive dictionary (e.g. as an aid in language learning).

A main reason why the discussion of description versus prescription up till now has been less than productive and that it has, for example, led to faith-related controversies, is to be found in the fact that only strong polyfunctional dictionaries had been targeted for investigation. It might be that for small languages and for many specialised languages only polyfunctional dictionaries could have been compiled, because the market would not allow the full spectrum of possible and necessary monolingual dictionaries. But this does not change the fact that such an argument does not apply to electronic dictionaries, where monofunctional dictionaries could be extracted from a huge mutual databank.

References

- Atkins, B.T.S. 1992-93. Theoretical Lexicography and its Relation to Dictionary-making. *Dictionaries. Journal of the Dictionary Society of North America* 14: 4-43.

- Bauer, Heinrich.** 1827–1833. *Vollständige Grammatik der neuhochdeutschen Sprache. Bd. I 1827, Bd. II 1828, Bd. III 1830, Bd. IV 1832, Bd. V 1833.* Berlin: Reimer.
- Bergenholtz, Henning.** 1992. Lemmaselektion in zweisprachigen Wörterbüchern. Meder, Gregor and Andreas Dörner (Eds.). 1992. *Worte, Wörter, Wörterbücher: Lexikographische Beiträge zum Essener Linguistischen Kolloquium:* 49-65. Tübingen: Niemeyer.
- Bergenholtz, Henning.** 1994. Fachsprache und Gemeinsprache: Lemmaselektion im Fachwörterbuch. Schaefer, Burkhard and Henning Bergenholtz (Eds.). 1994. *Das Fachwörterbuch. Fachwissen und seine Repräsentation in Wörterbüchern:* 285-304. Tübingen: Narr.
- Bergenholtz, Henning.** 1998. Das Schlaue Buch. Vermittlung von Informationen für textbezogene und textunabhängige Fragestellungen. Zettersten, Arne, Jens Erik Mogensen and Viggo Hjørnager Pedersen (Eds.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography May 2–5, 1996 at the University of Copenhagen:* 93-110. Tübingen: Niemeyer.
- Bergenholtz, Henning.** 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Bergenholtz, Henning and Rufus H. Gouws.** 2006. How to Do Language Policy with Dictionaries. *Lexikos* 16: 13-45.
- Bergenholtz, Henning and Mia Johnsen.** 2007. Log Files Can and Should Be Prepared for a Functionalistic Approach. *Lexikos* 17: 1-20.
- Bergenholtz, Henning and Uwe Kaufmann.** 1996. Enzyklopädische Informationen in Wörterbüchern. Weber, N. (Ed.). 1996. *Semantik, Lexikographie und Computeranwendungen:* 168-182. Tübingen: Niemeyer.
- Bergenholtz, Henning and Burkhard Schaefer.** 1977. Deskriptive Lexikographie. *Zeitschrift für germanistische Linguistik* 5(1): 2-33.
- Changuion, A.N.E.** 1844. Proeve van Kaapsch Taaleigen. Changuion, A.N.E. 1844. *De Nederduitsche Taal in Zuid-Afrika hersteld.* Cape Town: J.H. Collard.
- Engel, Eduard.** 1918. *Entwelchung. Verdeutschungswörterbuch für Amt, Schule, Haus, Leben.* Leipzig: Hesse & Becker.
- Fabricius, Johann Albert.** 1723. Usus imperium sine exceptione assertum oder die unumschränkte Gewalt des Gebrauchs/ gleichwie in andern Sprachen/ also auch in der Teutschen, behauptet wider die weit aussehenden Neuerungen der SprachLehrenden Malcontenten. C.F. Reichmanns *Poesie der Nieder-Sachsen. Zweyter Theil. Dem vorgedruckt einigen zwischen Herrn Doct. Fabricius und Herrn Prof. Richey freundlich-gewechselte Streit-Schriften über verschiedene, die Teutsche Sprache betreffende, Puncte:* 12-26. Hamburg: Johann Christoph Kießner.
- Gouws, Rufus H. and Liezl Potgieter.** 2010. Does Johnson's Prescriptive Approach Still Have a Role to Play in Modern-Day Dictionaries? *Lexikos* 20: 234-247.
- Greimas, Algirdas Julien and Joseph Courtes.** 1979. *Sémiotique: Dictionnaire raisonné de la théorie du langage.* Paris: Hachette.
- Haiman, John.** 1980. Dictionaries and Encyclopedias. *Lingua* 50: 329-357.
- Haßler, Gerda.** 2009. Normierung. Haßler, Gerda and Cordula Neis (Eds.). 2009. *Lexikon sprachtheoretischer Grundbegriffe des 17. und 18. Jahrhunderts:* 674-718. Berlin/New York: de Gruyter.
- Johnson, Samuel.** 1747. *The Plan of a Dictionary of the English Language.* Facsimile edition 1970. Menston: The Scholar Press.

- Johnson, Samuel.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton, T. & T. Longman et al.
- La Forêt, Gregorius.** 1760. *Lingva Germanica in Ore Danico. Det er: En nye Anviisning hvorledes en dansk, der vil tale Tydsk, har adskillige DANISMOS at tage sig vare for, hvorved der tillige handles grundig og udførlig om De Tydske Substantivorum Genere, tilstrækkelig om Declination og Conjugation og noget om Ordenes Sammensættelse. Forfattet af La Forêt. Og paa nye oplagt som et Tillæg til Schallers Grammatica*. Copenhagen: Andreas Hartvig Godiche.
- Longman Exams Dictionary*. 2006. Essex: Pearson Education.
- Richey, Michael.** 1723. *Usus Tyrannus precario imperans, oder Vernunft-mäßige Sätze von der Gewalt des Gebrauches in den Sprachen/insonderheit in der Teutschen, ingleichen ob und wie weit man demselben durch Regeln der SprachKunst Einhalt thun könne. C.F. Reichmanns Poesie der Nieder-Sachsen. Zweyter Theil. Dem vorgedruckt einigen zwischen Herrn Doct. Fabricius und Herrn Prof. Richey freundlich-gewechselte Streit-Schriften über verschiedene, die Teutsche Sprache betreffende, Puncte: 2-11*. Hamburg: Johann Christoph Kießner.
- Tarp, Sven and Rufus H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- The Danish Internet Dictionary* = Henning Bergenholtz in collaboration with Filip Bodilsen <<http://www.asb.dk/stafflist.aspx?pid=17064&focus=18766>>, Kathrine Carstensen <<http://www.asb.dk/stafflist.aspx?pid=17064&focus=21249>>, Kathrine Brosbøl Eriksen, Stine Busk Hedegaard <<http://www.asb.dk/stafflist.aspx?pid=17064&focus=18767>>, Helene H. Jensen, Emilie Dittmer Laursen, Katja Å. Laursen <<http://www.asb.dk/stafflist.aspx?pid=17064&focus=18768>>, Heidi Agerbo Pedersen <<http://www.asb.dk/stafflist.aspx?pid=17064&focus=18765>>. Database: Richard Almind: Den Danske Netordbog. Odense: Ordbogen.com 2010 <<http://www.ordbogen.com/>>(Fifth Edition).
- Wierzbicka, Anna.** 1993. What Are the Uses of Theoretical Lexicography? *Dictionaries. Journal of the Dictionary Society of North America* 14: 44-78.
- Woordeboek van die Afrikaanse Taal*. Stellenbosch: Buro van die WAT.
- Wustmann, G.** 1912. *Allerhand Sprachdummheiten. Kleine deutsche Grammatik des Zweiflhaften, des Falschen und des Häßlichen. Ein Hilfsbuch für alle die sich öffentlich der deutschen Sprache bedienen*. Strasbourg: Trübner.

Kontekstuele bepalers by die beplanning van skoolwoordeboeke

Herman L. Beyer, *Departement Taal- en Literatuurstudie, Universiteit van Namibië, Windhoek, Namibië (hbeyer@unam.na)*

Opsomming: Die doel van hierdie artikel is om tot die metaleksikografiese gesprek oor skoolwoordeboeke as 'n tipe pedagogiese woordeboek by te dra. Die fokus is op kontekstuele bepalers wat by die beplanning van skoolwoordeboeke verreken behoort te word: Binne die konteks van 'n veeltalige Namibië word enkele onderwyskundige sisteme verken wat 'n impak op die ontwerp van gedrukte skoolwoordeboeke vir die sekondêre skoolfase kan hê, naamlik skoolkurrikula, die nasionale taalbeleid vir skole en die beleid van leerdergerigte onderwys. Hierdie artikel voer ook aan dat die voor die hand liggende teikengebruiker van 'n skoolwoordeboek, naamlik die leerder, nie die enigste teikengebruiker is nie: Die onderwyser word as medeteikengebruiker gestel wanneer dit by die gebruik van woordeboeke in die onderriglersituasie kom. Verskillende tipes gevalle van woordeboekgebruik in die mikrokonteks van die onderriglersituasie word onderskei, naamlik *woordeboekonderrig*, *woordeboekgeïntegreerde taalonderrig* en *woordeboekgeondersteunde taalonderrig*. Hierdie tipes gebruik behoort komponente van 'n woordeboekpedagogie te wees, wat uiteindelik behoort te lei tot *outonome woordeboekgebruik* en die gepaardgaande vestiging van 'n *individuele woordeboekkultuur* by die leerder. Die leksikograaf moet al hierdie aspekte inreken wanneer 'n skoolwoordeboek beplan word ten einde die nodige leksikografiese vernuwings in te voer. Enkele voorstelle vir sodanige vernuwings en vir verdere navorsing word ten slotte aangebied.

Sleutelwoorde: GEBRUIKERSGERIGTE LEKSIKOGRAFIE, KONTEKS, LEERDER, LEERDERGERIGTE ONDERWYS, LEERWOORDEBOEK, LEKSIKOGRAAF, METALEKSIKOGRAFIE, NASLAANWOORDEBOEK, ONDERWYSER, PEDAGOGIESE WOORDEBOEK, SISTEEM, SKOOLWOORDEBOEK, TAALBELEID, TEIKENGEBRUIKER, WOORDEBOEKGEBRUIK, WOORDEBOEKGEÏNTEGREERDE TAALONDERRIG, WOORDEBOEKONDERSTEUNDE TAALONDERRIG, WOORDEBOEKONDERRIG, WOORDEBOEKPEDAGOGIE

Abstract: Contextual Determiners in the Planning of School Dictionaries.

The aim of this article is to contribute to the metalexical discussion of school dictionaries. The focus is on contextual determiners that should be accounted for in the planning of school dictionaries: Within the context of a multilingual Namibia, some educational systems that could have an influence on the design of printed school dictionaries for the secondary school phase, i.e. school curricula, the national language policy for schools and the policy of learner-centred education, are explored. This article also argues that the obvious target user of a school dictionary, i.e. the learner, is not the only target user: The teacher is seen as co-target user when dictionary use takes place in the teaching learning situation. Different types of dictionary use instances can be distinguished in the micro-context of the teaching learning situation, i.e. *dictionary skills teaching*, *dictionary-integrated language learning* and *dictionary-assisted language learning*. These types of use should form components of a dictionary pedagogy, which should ultimately result in *autonomous dictionary use* and the concurrent establishment of an *individual dictionary culture* with the learner. The lexicographer should take cognisance of all these aspects when a school dictionary is planned in order to intro-

duce the necessary lexicographical innovations. In closing, a few such innovations as well as suggestions for further research are proposed.

Keywords: CONTEXT, DICTIONARY PEDAGOGY, DICTIONARY SKILLS TEACHING, DICTIONARY USE, DICTIONARY-ASSISTED LANGUAGE LEARNING, DICTIONARY-INTEGRATED LANGUAGE LEARNING, LANGUAGE POLICY, LEARNER, LEARNER-CENTRED EDUCATION, LEARNING DICTIONARY, LEXICOGRAPHER, METALEXICOGRAPHY, PEDAGOGICAL DICTIONARY, REFERENCE DICTIONARY, SCHOOL DICTIONARY, SYSTEM, TARGET USER, TEACHER, USER-DRIVEN LEXICOGRAPHY

1. Inleiding

Die doel van hierdie artikel is om by te dra tot die metaleksikografiese gesprek oor skoolwoordeboeke. Die subklas skoolwoordeboeke vorm saam met die subklas aanleerderwoordeboeke die klas pedagogiese woordeboeke. Volgens Gouws en Tarp (2008: 66) bestaan daar verwarring wat die gebruik van die terme *skoolwoordeboek* en *aanleerderwoordeboek* betref. Terwyl skoolwoordeboeke afgestem is op skoolleerders wie se moedertaal die leksikografies behandelde taal is, fokus aanleerderwoordeboeke op die behoeftes van die teikengebruiker wat 'n vreemde taal aanleer (Gouws en Prinsloo 2005: 51-52).¹ Die veronderstelling is dus dat skoolwoordeboeke by verstek eentalige verklarende woordeboeke is, terwyl aanleerderwoordeboeke eentalig verklarende woordeboeke, tweetalige woordeboeke of tweetalig verklarende (die sg. "bilingualised") woordeboeke kan wees. Hierdie toegedigde kenmerke van die twee tipes pedagogiese woordeboeke behoort egter nie rigied toegepas te word nie: Die kenmerke wat 'n spesifieke tipe (pedagogiese) woordeboek vertoon, moet in die eerste plek die produk wees van die woordeboek se werklike doel, naamlik die vervulling van die teikengebruiker se inligtingsbehoefte (Gouws 2006: 74). Die onderskeid van Prinsloo en Gouws (2005: 51-52) word in beginsel in hierdie artikel gehandhaaf, hoewel aangetoon sal word dat kontekstuele faktore 'n invloed op woordeboektipologie kan uitoefen.

Verreweg die meeste navorsing in die pedagogiese leksikografie tot op hede fokus op aanleerderwoordeboeke (Jackson 2002: 77, Bogaards 2003: 28, Kernerman 2007: 141-142), met skoolwoordeboeke wat byna glad nie figureer nie (vgl. egter vir Suid-Afrika: Lombard 1990, Louw 2004, Gouws 2006 en Hiles 2010). Een van die oorsake hiervan is waarskynlik, volgens Gouws en Prinsloo (2005: 51),

a false impression that they are easy to compile and merely require a cut and paste approach to extract them from bigger monolingual dictionaries.

So 'n benadering lei tot disfunksionele leksikografiese produkte wat in werklikheid groter taalvaardigheid van die teikengebruiker (die leerder) vereis as wat redelik in 'n gebruikersgedrewe leksikografie verwag kan word. Hartmann (2001: 75), wat dieselfde argument aanvoer, merk op dat die kreatiewe didak-

tiek wat kenmerkend is van kinderwoordeboeke, in skoolwoordeboeke ontbreek. Hierdie kritiek was tot onlangs toe ook op die Afrikaanse leksikografie van toepassing, hoewel onlangse verwickelinge belofte toon dat skoolwoordeboeke toenemend funksioneel beplan word, vgl. bv. *Nuwe Woordeboek Sonder Grense* (Gouws et al. 2004), *HAT Afrikaanse skoolwoordeboek* (Luther 2009 — voortaan *HAT Skoolwoordeboek*) en die *Longman Grondslagfasewoordeboek Afrikaans/Engels* (Gouws et al. 2010). Ten spyte van hierdie verwickelinge is daar steeds ruimte vir die uitbreiding van skoolwoordeboeke se potensiaal.

Hoewel die geïdentifiseerde funksies van 'n bepaalde woordeboek 'n sentrale rol in die vasstelling van die datakategorieë en -verspreidingstruktuur van die woordeboek speel, poog hierdie artikel om 'n bydrae te maak tot die uitbreiding van skoolwoordeboeke se potensiaal deur kontekstuele faktore wat by die beplanning van skoolwoordeboeke 'n rol kan speel binne die raamwerk van 'n kommunikatiewe woordeboekmodel (vgl. Beyer 2006) te verken. Binne hierdie model verteenwoordig onder meer die woordeboek (en dié se funksies, inhoud en strukture), die teikengebruiker, die gebruiksituasie en die konteks onderskeibare maar hoogs interafhanklike komponente. Met die term *konteks* word verwys na daardie versameling omgewingsveranderlikes wat normaalweg buite die invloedseer van die leksikograaf en die gebruiker in hulle onderskeie kapasiteite val, maar waarbinne én die leksikografiese proses én woordeboekgebruik plaasvind (Beyer 2006: 60). Binne 'n bepaalde konteks kan ongereguleerde subkontekste bestaan, bv. die veeltaligheid van 'n gemeenskap, sowel as gereguleerde subkontekste oftewel sisteme, bv. 'n nasionale grondwet. Suid-Afrika bestaan uit 'n versameling veeltalige samelewings, wat taalkontak en gevolglik taalverandering meebring. Behalwe vir die breër konteks van veeltaligheid vind taalonderrig ook binne bepaalde sisteme plaas. Die identifisering en verrekening van hierdie sisteme, wat as bepalers in die beplanning van gedrukte skoolwoordeboeke vir die sekondêre fase (graad 8 tot 12) geld, is die fokus van hierdie artikel.

Ter afsluiting word enkele voorstelle vir leksikografiese vernuwing in en navorsing oor skoolwoordeboeke na aanleiding van die geïdentifiseerde bepalers en bestaande funksies aan die hand gedoen.

2. Makrosistemiese bepalers

2.1 Skoolkurrikula

2.1.1 Taalvakke

Die makrosistemiese faktor wat die leksikograaf (in opdrag van die uitgewer) waarskynlik eerste by die beplanning van 'n skoolwoordeboek oorweeg, is die relevante (taal)vaksillabusse. Met betrekking tot die Suid-Afrikaanse konteks meld Van der Merwe (2009: 301) dat skoolsillabusse sedert 2002 spesifieke "uitkomstes" ten opsigte van woordeboekgebruik stel, byvoorbeeld dat leerders

woorde in hulle alfabetiese plek sal kan vind en die spelling en betekenis van woorde sal kan nagaan.

Wat die Namibiese konteks betref, word daar in die sillabus van die junior primêre fase (graad 1 tot 4) geen eksplisiete melding van woordeboeke gemaak nie; daar word wel genoem dat leerders toegang tot 'n wye verskeidenheid leesstof behoort te hê (Ministry of Basic Education, Sport and Culture 2004: 2), wat uiteraard woordeboeke kan insluit. Die sillabusgids vir eerstetaalonderwysers in dié fase (Ministry of Education 2005) bevat egter 'n aanbevole jaarplan vir elke graad waarin woordeboeke baie duidelik figureer. In Tabel 1 hieronder word die relevante basiese vaardighede weergegee:

Tabel 1: Basiese vaardighede met betrekking tot woordeboeke in die junior primêre fase (vertaal uit Ministry of Education 2005)

| Graad | Tema | Basiese vaardigheid |
|-------|---|--|
| 1 | 3: Kultuur in die familie | Maak prentewoordeboeke (woorde wat met dieselfde letter begin). |
| 1 | 11: Veiligheid in en om die huis | Begin eie woordeboeke maak (dik potlode). |
| 1 | 12: Voedsel | Gaan voort met woordeboeke. |
| 1 | 13: Plaaslike plante | Gaan voort met woordeboeke. |
| 2 | 4: Plaaslike kultuur | Vind woorde in 'n woordeboek. |
| 2 | 13: Plaaslike voedsel | Skryf nuwe woorde in 'n selfgemaakte woordeboek. |
| 2 | 19: Water | Skryf alle nuwe woorde neer in die woordeboek. |
| 3 | 6: Openbare vakansiedae | Gebruik 'n woordeboek om die betekenis van woorde te vind. |
| 3 | 18: Identifikasie (van plante) | Beantwoord vrae skriftelik in gepaste taal en korrigeer spelling met behulp van 'n woordeboek of die onderwyser. |
| 3 | 19: Interafhanklikheid van diere en hulle habitat | Soek "diere"-woorde in 'n woordeboek op en beskryf hulle. |
| 4 | 10: Hoe boere ons mense voed | Skryf 'n kort verhaal oor 'n plaas (± 150 woorde). Kontroleer spelling in die woordeboek. |
| 4 | 23: Gesonde voeding | Vind die betekenis van nuwe woorde in 'n woordeboek. |
| 4 | 29: Stowwe en hulle eienskappe | Oefen daaglik korrekte spelling deur woordeboeke te gebruik. |
| 4 | 33: Klank | Lys nuwe woorde en vind hulle betekenis in konteks deur 'n woordeboek te gebruik. |

In die sillabus vir Afrikaans Eerste Taal in die senior primêre fase (graad 5 tot 7) word slegs verwag dat leerders hulle kennis van alfabetiese volgorde behoort te kan toepas "by die gebruik van verskillende soorte media, bv. woordeboeke, indekse, ens." (Ministry of Basic Education and Culture 2000: 17). Die sillabus vir Afrikaans Eerste Taal in die junior sekondêre fase (graad 8 tot 10) gee meer uitgebreide aandag aan woordeboeke met die formulering van die volgende basiese vaardighede onder verskillende doelstellings (Ministry of Education 2007: 9, 16). Leerders behoort naamlik in staat gestel te word om

- kort tekste van verskillende aard (storie, drama, koerantberig, woordeboekinskrywing, afkondiging, ens.) vlot en akkuraat voor te lees [slegs graad 8 — HLB],
- betekenis, spelling en idiomatiese gebruik van woorde in woordeboeke na te slaan,
- woordeboeke en ander beskikbare naslaanwerke te gebruik om spelling, betekenis en addisionele inligting na te slaan, soos afkortings, idiomatiese uitdrukkings, woordsoort, ens., en
- elektroniese bronne (waar beskikbaar) te gebruik vir naslaandoeleindes.

In die sillabusse vir Afrikaans Eerste Taal in die senior sekondêre fase (graad 11 en 12) word vir beide die gewone en hoër vlakke bloot van leerders verwag om "woordeboeke vir verskillende doeleindes" te kan raadpleeg (Ministry of Education 2006: 8, Ministry of Education 2006a: 8).

Dit is belangrik vir die leksikograaf om te beseft dat kurrikuleerders met betrekking tot woordeboeke slegs sodanige gebruiksvaardighede kan voorskryf as wat die beskikbare (skool)woordeboeke inligting bied. Indien die beskikbare woordeboeke byvoorbeeld nie aanduidings van antonieme bevat nie, kan sillabusse nie vereis dat leerders in staat behoort te wees om antonieme in woordeboeke te vind nie. Met betrekking tot skoolwoordeboeke kan daar dus 'n dooie kringloop ontstaan: Leksikograwe en uitgewers bepaal die datakategorieë wat in hulle woordeboeke aangebied sal word op grond van wat sillabusse ten opsigte van woordeboekgebruik voorskryf, en kurrikuleerders bepaal wat sillabusse ten opsigte van woordeboekgebruik sal voorskryf op grond van watter inligting in woordeboeke aangebied word. Die onus lê op die leksikografie om (a) skoolwoordeboeke te produseer wat nie bloot aan die minimum sillabusvereistes voldoen nie, maar wat binne die beperkinge van die beskikbare media ook innoverend is, en (b) aan belanghebbendes in die onderwyssektor leksikografiese voorligting te bied.

2.1.2 Basiese Inligtingkunde

Verwysings na woordeboekgebruik in Namibiese skoolkurrikula is nie tot die taalvakke beperk nie. In die sillabus vir die vak Basiese Inligtingkunde wat in

die senior primêre fase vir alle leerders aangebied word, kom woordeboeke as tema in graad 5 voor met die leerdoelstelling dat leerders moet leer hoe om 'n woordeboek te gebruik (Ministry of Education 2007: 11). Die basiese vaardighede is (ongelukkig) beperk tot leerders se vermoë om inligtings- en kommunikasietegnologieterne in 'n woordeboek te vind en te gebruik. Woordeboeke figureer nie in die leerinhoud vir graad 6 en 7 nie.

Wanneer die Namibiese sillabusse in chronologiese volgorde beskou word, wil dit voorkom asof daar 'n toenemende bewustheid onder besluitnemers oor die gebruik van woordeboeke en woordeboekonderrig aan die ontwikkel is. Sistematiese progressie in woordeboekvaardigheid ontbreek egter binne en tussen skoolfasies. Ten opsigte hiervan kan die leksikografie kurrikuleerders bystaan in die ontwikkeling van 'n deurlopende woordeboekpedagogie.

Die oorweging van kontekstueel-sistemiese faktore buiten kurrikula kan ook waarde tot skoolwoordeboeke toevoeg. Enkele sodanige faktore word vervolgens kortliks uiteengesit.

2.2 Die nasionale taalbeleid vir skole

In die veeltalige konteks van Namibië geld die grondwetlike bepaling dat Engels die (enigste) amptelike taal is en dat daarnaas dertien nasionale tale erken word. Volgens die nasionale taalbeleid vir skole (Ministry of Education and Culture 1993) word moedertaalonderrig tot in graad 3 toegelaat, met Engels wat as onderrigmedium vanaf graad 4 ingefaseer word en vanaf graad 8 as die enigste onderrigmedium in Namibiese staatskole geld. Die nasionale tale wat as skoolvakke aangebied word, word uiteraard deur die medium van die betrokke taal onderrig. Die implikasie is dat byvoorbeeld moedertaalsprekers van Afrikaans na graad 3 vir maksimaal 15,4% van hulle klastyd aan Afrikaans blootgestel word (vgl. Ministry of Basic Education, Sport and Culture 1996: 21). In alle ander vakke vind alle kommunikasie deur die medium van Engels plaas, wat groot druk op Afrikaanstalige leerders se moedertaalleksikon- en taalvaardigheidsontwikkeling plaas, aangesien hulle Engelse leksika in sekere gevalle vinniger uitbrei as dié van hulle moedertaal. In sommige gevalle sal leerders met Afrikaans as moedertaal byvoorbeeld 'n saak of begrip beter in Engels kan benoem as in Afrikaans. Indien van hierdie leerders dan verwag word om hulle deurlopend in Standaardafrikaans uit te druk, soos wat in die Afrikaansklas die geval is, kan dit gebeur dat die leerders 'n tweetalige woordeboek soos die *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary* (Louw 2007) raadpleeg om 'n bepaalde moedertaalekwivalent van 'n Engelse leksikale item (terug) te vind. In hierdie geval word dié tweetalige woordeboek dan tipologies gesproke nie as aanleerderwoordeboek gebruik nie, maar tereg as skoolwoordeboek — met die moedertaal as doeltaal. Binne die huidige Namibiese konteks is dit dus nie vergesog dat die subklas skoolwoordeboeke tipologies gesproke nié sonder meer deur eentalig verklarende

woordeboeke verteenwoordig word nie. Die veeltalige konteks en die geldende nasionale taalbeleid vir skole in Namibië het dus duidelik 'n invloed op die benutting van (skool)woordeboeke, wat spesifieke leksikografies relevante behoeftes weerspieël waarvolgens die relevante woordeboektipologieë bepaal behoort te word.

2.3 Die beleid van leerdergerigte onderwys

'n Makrosistemiese faktor wat 'n gulde geleentheid vir die gebruik van skoolwoordeboeke bied, maar wat ook by die ontwerp van skoolwoordeboeke in ag geneem behoort te word, is die beleid van *leerdergerigte onderwys*, wat voortspruit uit sosiale konstruktivisme (Combrink et al. 1997: 6). Volgens die *Pilot Curriculum Guide for Formal Basic Education* (Ministry of Basic Education, Sport and Culture 1996: 25) behoort leerdergerigte onderwys die volgende eienskappe te vertoon:

- the starting point at each stage of a learning process is each learner's existing knowledge, skills, interests and understanding, derived from previous experience in and out of school;
- the natural curiosity and eagerness of all young people to learn to investigate and make sense of a widening world must be catered for by a variety of challenging and meaningful tasks;
- the learner's perspective must be appreciated and taken into consideration in the work of the school;
- learners should be empowered to think and take responsibility not only for their own, but for each other's learning and development; and
- they should be involved as partners in, rather than receivers of, education.

In die besprekingsdokument *Learner-Centred Education in the Namibian Context: A Conceptual Framework* (NIED 2003: 25-27) word die implikasies van leerdergerigte onderwys op leer- en ondersteuningsmateriaal uitgelig. Die volgende punte wat op skoolwoordeboeke as leer- en ondersteuningsmateriaal van toepassing kan wees, is geëkstraheer:

- Die herwaarding van die rol van leer- en ondersteuningsmateriaal in die bereiking van substansiële leer: Materiaal moet die leerproses ondersteun in plaas van domineer.
- Leerders se vorige ervaring en leer moet in ag geneem word en geleenthede vir verkenning en probleemoplossing bied.
- Materiaal behoort 'n raamwerk vir die fasilitering van leer te skep waarin vraagstelling, vergelyking en verkenning deurgaans op verworwe kennis bou.

Die innoverende gebruik van skoolwoordeboeke in die taalklas kan effektief bydra tot die toepassing van leerdergerigte (taal)onderrig, mits skoolwoordeboeke voldoende didaktiese elemente bevat en taalonderwysers voldoende in hierdie verband opgelei is. Dít vra om leksikografiese vernuwing en gerigte woordeboekpedagogieë.

2.4 Kruiskurrikulêre onderrig

Binne die beleid van leerdergerigte onderwys figureer 'n belangrike benadering, naamlik dié van *kruiskurrikulêre onderrig*. Hierdie benadering word soos volg in 'n beleidsdokument toegelig (Ministry of Basic Education and Culture 1998: 8):

Cross-curricular teaching can be formally organised around topics and themes. A common topic in two or more subjects is planned so that it will be taught simultaneously, e.g. population statistics. The Mathematics teacher might go into the technical problems of population statistics and how that could be presented, the teacher of Geography will take up the significance of population statistics for understanding population growth, demographic aspects and the relation of population and land resources. The English teacher may present literary extracts illustrating the human dilemmas which result from overpopulation.

Die opdrag aan die Namibiese Afrikaansonderwyser binne hierdie benadering is duidelik: Ontwikkel leerders se taalvaardigheid in Afrikaans parallel met die betrokke sentrale tema wat eksklusief deur die medium van Engels in die ander vakke behandel word. Hierdie opdrag sluit vanselfsprekend die parallelle uitbouing van leerders se Afrikaanse leksika in. Die gebruik van 'n tweetalige (skool)woordeboek om moedertaalekwivalente vir Engelse terme wat in al die ander vakke gebruik word te vind, dit wil sê die gebruik van sodanige woordeboek as *skoolwoordeboek* in plaas van *aanleerderwoordeboek*, is gevolglik binne 'n leerdergerigte situasie nie 'n onredelike veronderstelling nie. Trouens, dit kan selfs 'n staande opdrag in die Afrikaansklas wees: Aan die begin van elke Afrikaansperiode word 'n paar minute afgestaan aan die vind van Afrikaanse vertaalekwivalente vir Engelse terme wat in ander vakke gebruik is deur hulle in 'n tweetalige skoolwoordeboek na te slaan. So 'n klasbeleid kan ook tot die vestiging van 'n taalbewustheid onder leerders bydra.

Die benadering van kruiskurrikulêre onderrig binne die Namibiese konteks en die leksikografies relevante inligtingsbehoefte wat daaruit voortspruit, kan dus ook 'n hersiening van bestaande woordeboektipologieë (m.b.t. pedagogiese woordeboeke) motiveer.

2.5 Perspektief op makrosistemiese bepalers

In die voorafgaande bespreking is enkele makrosistemiese faktore uitgelig wat die beplanning en gebruik van skoolwoordeboeke kan — en behoort te — beïn-

vloed. Daar is nie na 'n uitputtende identifikasie en beskrywing van die tersaaklike faktore gestreef nie. Die sentrale argument is wel dat skoolwoordeboeke wat slegs op sillabusvereistes met betrekking tot woordeboekgebruik gegrond is nie die potensiaal van die pedagogiese leksikografie ten opsigte van hierdie woordeboektype sal verwesenlik nie. Presies hoe die oorweging van die bespreekte faktore tot beter skoolwoordeboeke aanleiding kan gee, sal later in die artikel aan die hand van enkele voorstelle vir leksikografiese vernuwing toegelig word.

3. Die teikengebruiker en mikrosistemiese bepalers

In die moderne leksikografie staan die gebruikersperspektief voorop: Die inligtingsbehoefte en naslaanvaardighede van 'n goed gedefinieerde teikengebruiker is kritiese veranderlikes in die beplanning van 'n woordeboek (vgl. o.m. Gouws en Prinsloo 2005: 39). Op die vraag wie die teikengebruiker van 'n skoolwoordeboek is, is die vanselfsprekende antwoord "die leerder". 'n Leerder ondervind binne (of buite) die onderrigleersituasie 'n behoefte aan sekere inligting wat deur die kommunikatiewe en kognitiewe funksies van die gepaste woordeboek bevredig kan word, indien die leerder se naslaanvaardighede hom/haar toegang tot die relevante data bied (vgl. Tarp 2008). Die leksikograaf wat 'n akkurate profiel van die leerder kan saamstel, kan 'n funksionele skoolwoordeboek tot stand bring.

Volgens die *Dictionary of Lexicography* (Hartmann en James 1998: 107) is 'n pedagogiese woordeboek "a reference work specifically designed for the practical didactic needs of teachers and learners of a language". Hierdie definisie maak die belangrike punt dat 'n skoolwoordeboek ook *die praktiese didaktiese behoeftes van die taalonderwyser* behoort aan te spreek: Die taalonderwyser is medeteikengebruiker en mag nie in die beplanning van 'n skoolwoordeboek oorgesien word nie. Hiervoor moet die leksikograaf inligting oor die onderrigleersituasie as mikrosistemiese konteks faktor inwin.

3.1 Die onderrigleersituasie

In waarskynlik die meeste gevalle word woordeboekgebruik in die onderrigleersituasie op die een of ander manier deur die onderwyser gestimuleer. Hierdie stimulering kan in die volgende gevalle plaasvind: (a) Die onderwyser bied onderrig in woordeboekgebruik aan, wat in die ideale situasie die praktiese toepassing van woordeboekgebruik deur die leerders insluit. Vir hierdie doel voorsien baie uitgewers ook woordeboekwerkboeke en -gidse, bv. *HAT Aktiwiteite en speletjies by die Afrikaanse skoolwoordeboek en Afrikaanse sakwoordeboek* (Luther 2010). (b) In die uitvoering van 'n taak (bv. die skryf van 'n opstel in die klas) ondervind die leerder 'n behoefte aan bepaalde leksikografies relevante inligting. Die leerder nader die onderwyser of 'n medeleerder om hulp,

wat óf die nodige inligting sonder meer aan die leerder verskaf ('n nie-ideale, nie-leerdergerigte situasie), óf die leerder na die woordeboek verwys (die ideale, leerdergerigte situasie: stimulering van woordeboekgebruik). Indien die leerder effektiewe woordeboekonderrig ondergaan het en 'n naslaankultuur word in 'n leerdergerigte onderrigleersituasie aangemoedig, sal hy/sy die woordeboek direk nader. Gouws (2009: 83) verwys na hierdie tipe woordeboekgebruik as geïsoleerde woordeboekgebruik. (c) Die taalonderwyser integreer die woordeboek op voorafbeplande wyse in die onderrigleersituasie deur leerders woordeboeke te laat gebruik om by te dra tot die bereiking van die een of ander lesdoelwit. Geval (a) staan bekend as *woordeboekonderrig* (WO — as komponent van 'n woordeboekpedagogie) en is gewoonlik afgestem op die verwerwing van daardie basiese vaardighede in die sillabus wat woordeboekgebruik spesifiseer, soos in 2.1 aangetoon. Geval (b) kan, na analogie van die term *rekenaargesteurde taalonderrig*, beskryf word as *woordeboekondersteunde taalonderrig* (WST). Geval (c) kan beskryf word as *woordeboekgeïntegreerde taalonderrig* (WIT), wat Gouws (2009: 83) ook *geïntegreerde woordeboekgebruik* noem. Die doelwit van elkeen van hierdie gevalle van woordeboekgebruik kan soos volg uiteengesit word:

WO: die verwerwing van 'n bepaalde woordeboekvaardigheid,

WST: die inwinning van die benodigde inligting om 'n taak te voltooi, en

WIT: die bereiking van 'n lesdoelwit deur leerdergerigte onderrigleeraktiwiteite.

'n Verdere belangrike verskil tussen WST enersyds en WO en WIT andersyds is dat WST in die ideale situasie deur die leerder geïnisieer word, terwyl die inisieerder by WO en WIT die taalonderwyser is. Die kollektiewe algemene doelstelling van al drie gevalle is die bereiking van 'n situasie waarin die leerder 'n woordeboekgeletterde gebruiker is wat *outonome woordeboekgebruik* (OWG) toepas en sodoende 'n *individuele woordeboekcultuur* (IWK) besit. Ongelukkig is dit in baie onderrigleersituasies die veronderstelling — selfs die verwagting — dat WST alleen tot OWG sal lei. In die uitsluitlik WST-situasie probeer woordeboekgeletterde leerders om woordeboeke te gebruik, wat aanleiding gee tot onsuksesvolle naslaanprosedures en die gepaardgaande ongunstige houding jeens woordeboekgebruik (vgl. Carstens 1995: 107-109). 'n Skoolwoordeboekpedagogie behoort al hierdie gevalle van woordeboekgebruik te inkorporeer, en hoewel die tipes afwisselend kan voorkom, is die volgende oorkoepelende toepassingsorde pedagogies gewens, met die mikrokonteks van die onderrigleersituasie wat deur die stippellyn afgebaken word:

{ WO → WIT → WST } → OWG

Dit is belangrik om in gedagte te hou dat die algemene doelstelling van 'n woordeboekpedagogie nie bloot OWG is nie, maar dat dit progressief tot 'n

IWK behoort te lei. Selfs die vestiging van 'n IWK behoort nie die hoofdoel te wees nie, want, soos Béjoint (1989: 209) beklemtoon,

dictionary use is not an end in itself; it is only a means that can be used to improve one's mastery of the language.

Dit word dus duidelik dat WO, soos in skoolsillabusse voorgeskryf, maar die eerste (noodsaaklike) stap tot OWG en uiteindelik 'n IWK is. Die komplementerende benadering van WIT dra daartoe by dat woordeboeke hulle volle potensiaal as instrumente in taalonderrig oftewel taalverwerwing kan verwesenlik.

3.2 Woordeboekgeïntegreerde taalonderrig (WIT)

Om WIT verder van WST te onderskei, kan 'n kommunikasieketting ten opsigte van die leer(onderrig)aktiwiteit in elke benadering beskou word:²

WST: ... → leerder (→ taalonderwyser → leerder) → WOORDEBOEK → ...

WIT: ... → taalonderwyser → leerder → WOORDEBOEK → ...

Soos reeds aangetoon, wys hierdie uiteensetting ook op die inisieerder van woordeboekgebruik in elke geval. In die WST-situasie ondervind die leerder onafhanklik 'n behoefte aan leksikografies relevante inligting en inisier (uiteindelik) woordeboekgebruik. Die inligting wat die leerder inwin, word benut om 'n taak te voltooi. In die WIT-situasie inisier die taalonderwyser woordeboekgebruik deur by die leerders 'n behoefte aan leksikografies relevante inligting te skep en die geleentheid te bied om 'n woordeboek gerig te gebruik. Ter illustrasie volg 'n eenvoudige voorbeeld van WIT:

Die taalonderwyser wil die volgende basiese vaardigheid in die sillabus vir Afrikaans Eerste Taal in graad 10 op 'n leerdergerigte manier vestig of konsolideer (Ministry of Education 2007: 13): Leerders behoort in staat te wees om

- hoofletters en leestekens korrek en effektief te gebruik (kappie, koppelteken, deelteken, afkappingsteken en aksentteken).

Die fokus is op die gebruik van die afkappingsteken by meervouds- en verkleinwoordvorme. Die onderwyser identifiseer die volgende relevante spelreëls volgens die leerstof of, soos in hierdie geval, die *Afrikaanse woordelys en spelreëls* (Taalkommissie 2009: 7-11):

Reël 2.1: Die afkappingsteken word gebruik by meervoudsvorme op **-s** en die verkleiningsvorme van naamwoorde wat op die vokaalletters **i**, **o** en **u** eindig.

Reël 2.7: Die afkappingsteken word gebruik by meervoudsvorme op **-s** en verkleiningsvorme van naamwoorde wat eindig op 'n lank uitgespreekte **a** ('n [a:]-klank wat met slegs een **a** gespel word) wat die hoofklem dra.

Reël 2.8: Die afkappingsteken word gebruik by die meervoudsvorme op **-e** en die verkleiningsvorme van selfstandige naamwoorde wat eindig op 'n **e** wat nie uitgespreek word nie.

Die volgende onderrigleeraktiwiteit, geskoei op die beginsels van leerdergerigte onderwys, word voorberei: Die leerders (wat reeds WO ontvang het) word gevra om 'n gegewe lys woorde in die woordeboek na te slaan en ten opsigte van elke woord (a) die meervouds- en verkleinwoordvorm na te gaan en neer te skryf, en (b) die vokaal of diftong wat volgens die woordeboek die hoofklem in elke woord dra, te onderstreep (soos in die *HAT Skoolwoordeboek*). Die lys woorde bestaan uit items wat aan die reëls hierbo onderhewig is sowel as items waarop dié reëls nie van toepassing is nie. Hierdie aktiwiteit verteenwoordig die moment van woordeboekintegrasie. Nadat die leerders die woordeboek gebruik het, word hulle opdrag gegee om die woorde volgens meervouds- en verkleinwoordvorming te klassifiseer en om in die klas woorde waarby die afkappingsteken voorkom uitspraakpatrone te soek. Die onderwyser lei die leerders dan gaandeweg om die tersaaklike reëls van die taalkundige gegewe af te lei in plaas daarvan om bloot die reëls weer te gee, met ander woorde 'n induktiewe in plaas van 'n deduktiewe werkswyse (vgl. Meij et al. 1985: 6) word gevolg waarin WIT bydra tot leerdergerigte taalonderrig. Ter ondersteuning van hierdie benadering het Meij et al. (1985: 26-27) met betrekking tot die onderrig van spelling bevind

dat leerlinge wat al die reëls loop leer het, skaars beter spellers is as die wat maar net goeie lesers is. Dus, ten spyte van die feit dat ons leerlinge spelreëls van jongs af leer, is die waarde daarvan nie baie duidelik nie.

Al meer en meer lyk dit of die interne reël wat elke gebruiker vir homself aflei, van meer waarde kan wees.

Hierdie voorbeeld van WIT verteenwoordig ook 'n geval waar die woordeboek as *leerwoordeboek* in plaas van *naslaanwoordeboek* (vgl. Hausmann 1977: 144) gebruik word, want die woordeboekgebruik behels nie 'n enkele naslaanprosedure wat ten doel het om spesifieke inligting met betrekking tot een bewerkingsseenheid (bv. 'n lemma) te onttrek nie, maar 'n reeks naslaanprosedures waardeur gelyksoortige inligting uit verskillende woordeboekartikels onttrek en daarna vir die doel van studie vergelyk word. In hierdie geval het die woordeboek naamlik 'n kognitiewe eerder as 'n kommunikatiewe funksie vervul (vgl. Tarp 2008: 46). Dit is hier belangrik om daarop te wys dat WIT nie die uitsluitlike gebruik van woordeboeke as leerwoordeboeke veronderstel nie. WIT sluit geensins die gebruik van woordeboeke as naslaanwoordeboeke uit nie; trouens, in die ideale situasie sou die leerder tydens WST en uiteindelik OWG ook die woordeboek as leerwoordeboek kon benut. 'n Leerdergerigte skoolwoordeboek wat wil bydra tot leerdergerigte onderwys, behoort as beide naslaanwoordeboek en leerwoordeboek gebruik te kan word, hoewel die ontwerp van 'n skoolwoordeboek nie leerwoordeboekfunksionaliteit ten koste van naslaanwoordeboekfunksionaliteit mag bevorder nie. Vanuit 'n leksikografiese

oogpunt kan leerwoordeboekfunktionaliteit as toegevoegde waarde beskou word. Die onderskeidende kenmerk van 'n woordeboek lê juis in sy naslaanfunktionaliteit (vgl. Tarp 2008: 46), hoewel hierdie kenmerk op sý beurt weer nie die moontlikheid van 'n uitgebreide funktionaliteit behoort uit te sluit nie. Dit is onder meer op hierdie punt waar die leksikograaf met vernuwing kan reageer.

Die breë konsep van WIT is nie nuut nie; trouens, die nasionale vakbeleid vir Engels Eerste Taal bevat die volgende stelling ten opsigte van woordeboekgebruik (Ministry of Education 2008: 4):

Teachers should plan lessons in such a way that learners are expected to use dictionaries.

'n Soortgelyke stelling kom egter nie in die nasionale vakbeleid vir Afrikaans Eerste Taal voor nie.

Baie taalonderwysers is nie bevoeg om WIT toe te pas nie, hoofsaaklik omdat hulle self nooit aan 'n woordeboekpedagogie blootgestel is nie. Opleiding in woordeboekgebruik moet in die eerste plek by onderwysers begin (Béjoint 1989: 208), en dié aspek moet in enige deeglike woordeboekpedagogie vir skole neerslag vind.

4. Voorstelle vir leksikografiese vernuwing en navorsing

Hierdie afdeling bied enkele voorstelle vir vernuwing in en navorsing oor skoolwoordeboeke, veral met die oog op leerdergerigte taalonderrig. Die fokus val op 'n aantal bestaande en potensiele datakategorieë in die sentrale lys, hoewel 'n uitgebreide dataverspreidingsstruktuur (d.w.s. 'n raamstruktuur) vir 'n skoolwoordeboek veronderstel word (vgl. Gouws en Prinsloo 2005: 58). Buitetekste word weens beperkte ruimte in hierdie artikel buite rekening gelaat. Voorbeelde van woordeboekartikels kom uit die *HAT Skoolwoordeboek* en hipotetiese woordeboekartikels is op artikels uit hierdie woordeboek gebaseer.

4.1 Die makrostruktuur

Binne die Namibiese konteks is die aangewese medium vir skoolwoordeboeke tans steeds die gedrukte medium. Beide die medium en die tipologiese klas waartoe die skoolwoordeboek behoort plaas sekere beperkinge op die omvang daarvan. 'n Skoolwoordeboek kan nie dieselfde omvang as 'n omvattende woordeboek hê nie, en daarom moet die leksikograaf 'n beperkte lemmalys saamstel. In die *HAT Skoolwoordeboek* word geen melding gemaak van die kriteria wat vir die samestelling van die makrostruktuur aangelê is nie. Op die flapteks word bloot genoem dat die woordeboek "meer as 30 000 woorde en hul verbuigings, duisende vaste uitdrukkings en idiome, inligting oor die herkoms van woorde en afkortings" bevat. In die Voorwoord van die *Oxford Secon-*

dary School Dictionary (Oxford University Press Southern Africa 2006: iv) word gemeld dat dié woordeboek, behalwe woorde uit Suid-Afrikaanse Engels, "the special vocabulary of curriculum subjects" bevat. Die veronderstelling is dat die uitgewers skoolsillabusse en leerstof raadpleeg en gespesialiseerde korpora en databasisse gebruik by die seleksie van lemmata. Indien korpora van leerders se skryfwerk beskikbaar is, sou 'n proskriptiewe benadering (vgl. Tarp en Gouws 2008: 239-240) oorweeg kon word. Woorde wat leerders geneig is om verkeerd te spel, kan byvoorbeeld geïdentifiseer word, en in hulle verkeerde spelling opgeneem en van die nodige proskriptiewe kommentaar voorsien word, soos in die volgende hipotetiese artikel:

~~parralel~~ Verkeerde spelling van **parallel**

Wat betref die toepassing van gebruiksfrekwensie met behulp van korpora om 'n makrostruktuur saam te stel, het De Schryver et al. (2006) bevind dat die mees frekwente woorde nie noodwendig dié woorde is wat woordeboekgebruikers opsoek nie — dit kan juis vir inligting oor die minder frekwente en derhalwe minder bekende woorde wees waarvoor 'n woordeboek geraadpleeg word. Afgesien van watter benadering tot die identifisering van lemmata gevolg word, is die uiteindelijke vrae vanuit 'n funksionele oogpunt: Watter woorde slaan leerders vir watter doel na? Watter woorde wil onderwysers hê moet leerders vir watter doel naslaan? Gebruik(er)navorsing om hierdie twee vrae te beantwoord, is dringend nodig. Verder moet die leksikograaf sy/haar beplande woordeboek se potensiele bydrae tot leerdergerigte onderwys in gedagte hou. Dit veronderstel dat die leksikograaf ook van taalonderrigmetodiek kennis moet neem.

4.2 Die mikrostruktuur

4.2.1 Woordsoorte

Die aanduiding van woordsoorte behoort uiteraard in ooreenstemming met die betrokke kurrikula te wees. Die hantering van die lemma **haar** in die *HAT Skoolwoordeboek* sien soos volg daaruit:

haar¹ *vnw.* (woord wat verwys na 'n vroulike persoon as dit nie die onderwerp van 'n sin is nie; iets wat 'n vrou s'n is, behoort aan *haar*, is *hare*): *Ek sien haar al van ver af aankom.* ◦ *Sy leen nooit haar boeke uit nie.* ◦ *Gee die teddiebeer vir jou sus-sie; dis hare.*

Dat daar geen eksplisiete onderskeid tussen die persoonlike en besitlike voor-naamwoord *haar* getref word nie, maak die artikel onbruikbaar vir WIT (én WST) wanneer dit by die onderrig van woordsoorte kom. In 'n aangepaste bewerking behoort die voorbeeldsinne ook op eksplisiete wyse deur 'n gepaste

leksikografiese prosedure aan die betrokke subwoordsoort gekoppel te word. Vergelyk die volgende voorgestelde herbewerking:

haar¹ *vnw.* ▣ *pers. vnw. wat verwys na 'n vr. pers. en wat nie die onderwerp van 'n sin is nie: Ek sien haar al van ver af aankom.* → **sy**³ ▣ *besitl. vnw. wat besit deur 'n vr. pers. aandui: Sy leen nooit haar boeke uit nie.*

In die bostaande artikel gaan die nie-tipografiese struktuurmerker ▣ die subwoordsoort vooraf om 'n afsonderlike soeksone af te baken waarbinne ook die relevante voorbeelde gegroepeer word. 'n Subwoordsoortaanduider hoef nie slegs 'n enkele afkorting te wees nie: In hierdie geval is die volledige subwoordsoortaanduider vir die persoonlike voornaamwoord *pers. vnw. wat verwys na 'n vr. pers. en wat nie die onderwerp van 'n sin is nie.* ('n Verwysing na die onderwerpsvorm van die voornaamwoord is ook ingevoeg.) Effektiewe WO en 'n funksionele gebruikersgids behoort te verseker dat die afkortings wat gebruik word, nie vir die hoërskoolleerder 'n probleem is nie. (Trouens, die gebruik van afkortings in woordeboekartikels bied die geleentheid vir WIT wanneer afkortings behandel word.)

4.2.2 Intensiewe vorme en vergelykings

In die *HAT Skoolwoordeboek* word intensiewe vorme (bv. *brandarm*) en vergelykings (bv. *so arm soos 'n kerkmuis*) in onderskeidelik 'n voorteks getiteld "Intensiewe vorme" en 'n agterteks getiteld "Vergelykings" in alfabetiese volgorde volgens die betrokke adjektief gelys. So kom intensiewe vorme vir 111 adjektiewe in die betrokke voorteks en 96 vergelykings in die betrokke agterteks voor. Hoewel die tersaaklike datakategorieë in enkele maklik toeganklike tekste saamgegroepeer is, bevorder dié tipe aanbod nie leerdergerigte taalonderrig of WIT nie. Vanuit 'n leerdergerigte oogpunt sal hierdie datakategorieë meer doeltreffend in die mikrostruktuur aangebied word,³ soos in die volgende hipotetiese artikel van die lemma **arm**²:

arm² *b.nw., bw.* [~ of ~e; **armer, die armste**] **1 arm wees** baie min geld en nie veel besittings hê nie. ◊ INTENS. VORM: *brandarm* ◊ VERG.: *kerkmuis*. **2 arm aan iets wees** nie veel daarvan hê/besit nie: 'n *skaam meisie, arm aan woorde* ◊ *grond wat arm aan stikstof is*. **3** ongelukkig; ellendig: *Die arme mense het sopnat gereën.*

Indien die taalonderwyser om die een of ander rede 'n lys van intensiewe vorme of vergelykings wil saamstel, kan dit met behulp van die woordeboek 'n opdrag aan die leerders wees — wat óók 'n geval van WIT verteenwoordig. Leerders kan verder gevra word om (hulle eie kreatiwiteit te gebruik om) vorendag te kom met intensiewe vorme en vergelykings van adjektiewe waarvoor daar nié sodanige data in die woordeboek verstrekkend word nie. WIT kan dus bou nie net op wat die woordeboek bied nie, maar ook op wat die woordeboek nié bied nie.

4.2.3 Metafore

Die aanduiding van metafore is 'n potensiële datakategorie vir 'n skoolwoordeboek. Vergelyk die volgende hipotetiese artikel:

morsig *b.nw., bw. [-e]* vuil; smerig: 'n morsige vent wat nooit bad nie ◦ Hoe kan 'n mens in so 'n morsige huis lewe? METAF.: vark (*neerh.*)

Weer eens sal WO en 'n funksionele gebruikersgids die leerder met die interpretasie van die inskrywings "METAF.: vark (*neerh.*)" help: "Om iemand (op neerhalende wyse) as morsig te beskryf, kan ook gesê word dat daardie persoon 'n vark is."

4.2.4 Idiome

Leerders wat idiome (en metafore) gepas in hulle skryfwerk gebruik, toon meer gevorderde taalbeheersing as diegene wat nie dié vaardigheid demonstreer nie. Gevolglik presteer sodanige leerders ook beter. Die leer van lyste idiome vir eksamen- of opstelskryfdoeleindes vanweë die moontlikheid dat hulle êrens nuttig te pas kan kom, getuig van 'n soortgelyke benadering as die uitwendige leer van spelreëls (vgl. 4.1). Sodanige benadering word bevorder deurdat woordeboeke tradisioneel met betrekking tot idiome slegs vir die funksie van teksresepsie voorsiening maak deur hulle opname in artikels van lemmata wat een van hulle komponente verteenwoordig. Die betrokke datakategorie is dus slegs via een soekroete toeganklik. Vergelyk die volgende voorbeeld:

onkant *b.nw. onkant wees* (*rugby, hokkie, sokker*) in 'n posisie wees waarin jy nie mag speel nie. ♦ **iem. onkant betrap** iemand in 'n situasie verras sodat hy/sy nie dadelik weet hoe om te reageer nie: *Sy huweliksaansoek het haar onkant betrap.*

Oor die leksikale en lemmastatus van idiome word nie uitgewei nie — vgl. hiervoor Gouws (1989: 98-101). Die woordeboekgeletterde leerder wat die idioom *iemand onkant betrap* tydens teksresepsie teëkom, kan dit aan die hand van die bostaande artikel interpreteer. Dieselfde leerder kan dit egter nie met behulp van die woordeboek as 'n gepaste idioom vind om 'n vorm van verrassing in 'n teksproduksiesituasie uit te druk nie. Hierdie leemte kan minstens gedeeltelik oorkom word deur 'n nuwe datakategorie in die artikel van die semanties verbandhoudende of sinonieme lemma in te voer, soos in die volgende hipotetiese artikel:

verras *ww. [het ~] iem. verras* iets doen wat iemand nie verwag het nie: *Die man het sy vrou met 'n bos blomme verras.* ◦ *Die vyand het hulle verras deur hulle van agter aan te val.* ◦ *Die Springbokke se oorwinning het baie mense verras.* ►IDM.: **onkant**

Die datakategorie wat deur die inskrywingsreeks "►IDM.: **onkant**" verteenwoordig word, verwys die leerder wat **verras** opsoek, na die artikel van die lemma **onkant** vir 'n idioom wat hy/sy sou kon gebruik om die begrip *verras* uit te druk sodat dié data dus politoeganklik en funksioneel gelykwaardig aan dié van leksikale lemmata gemaak word. Die behandeling van die idioom gee aan die leerder 'n aanduiding van die gepastheid daarvan vir die relevante diskoerskonteks. Alle idiome kan egter nie op hierdie manier politoeganklik gemaak word nie; daarom is verdere studie en voorstelle in hierdie verband nodig.

4.2.5 Afkortings

Indien kurrikula van leerders verwag om afkortings ook te kan gebruik en nie net te interpreteer nie, geld dieselfde argument as by idiome hierbo. Die afkortings van afkortbare lemmata behoort in die betrokke woordeboekartikels opgeneem te word, soos in die volgende hipotetiese artikel deur die inskrywingsreeks "AFK.: asb.":

as·se·blief *bw., tw.* [AFK.: asb.] 'n woord waarmee 'n mens iets hoflik vra: *Gee asseblief die suiker aan.* ◦ *Kom saam, asseblief.*

4.2.6 Ingevoegde binnetekste

Ingevoegde binnetekste is reeds 'n algemene verskynsel in skoolwoordeboeke. Vergelyk die volgende voorbeeld met die opskrif "Spelling" na die artikel van die lemma **lei**² in die *HAT Skoolwoordeboek*:

lei¹ en **lei**² is homonieme; **lei** en **ly** is homofone. Maak seker dat jy die verskil in spelling en betekenis ken.

Vanuit 'n leerdergerigte benadering sou betoog kon word vir inhoude wat nie bloot inligting voorsien nie, maar die leerder deur onder meer vraagstelling uitdaag om self ondersoek in te stel en te verken en sodoende 'n vennoot in die leerproses te word (vgl. 2.3). Ter ondersteuning hiervan sou die volgende herwerking van bostaande inhoud voorgestel kon word:

lei¹ en **lei**² is homonieme; **lei** en **ly** is homofone. Wat is 'n homoniem? Wat is 'n homofoon? Wat is die verskil in betekenis tussen *lei* en *ly*?

Deur die inhoud vernuwend in die vorm van vrae aan te bied, neem die woordeboek self deel aan die stimulering van 'n leksikografies relevante inligtingsbehoefte by die leerder, en word leerdergerigte leer sowel as woordeboekgebruik bevorder. (Die veronderstelling is uiteraard dat die antwoorde op sodanige vrae in die betrokke woordeboek te vinde is.)

4.3 Perspektief op leksikografiese vernuwing

Die elemente van vernuwing wat in die voorafgaande bespreking oorsigtelik aangebied is, verteenwoordig slegs enkele opsigte waarin skoolwoordeboeke vir 'n bepaalde konteks, soos dié van die onderwys in Namibië, funksioneel beplan kan word. Veral met die oog op leerdergerigte taalonderrig en WIT as 'n belangrike skakel in die proses wat tot OWG en uiteindelik 'n IWK kan lei, is leksikografiese vernuwing gebaseer op verdere navorsing nodig.

5. Slot

Die kennisname van sillabusvereistes en die bestudering van handboeke is maar sommige van die aspekte wat ter sprake is wanneer 'n skoolwoordeboek beplan word. By die bestudering van sillabusse behoort nie slegs op die vereistes met betrekking tot woordeboekgebruik gefokus te word nie, maar leksikograwe kan volledige sillabusse deurwerk en hulleself ten opsigte van elke doelstelling en doelwit afvra: Kan die skoolwoordeboek 'n datakategorie akkommodeer wat kan bydra tot die bereiking van hierdie doelstelling of doelwit?

Afgesien van sillabusse moet ander sistemiese faktore as die produkte van die konteks waarbinne woordeboekgebruik plaasvind, deeglik in skoolwoordeboekbeplanning verreken word. Hierdie eis veronderstel dat die leksikograaf in die beplanning van 'n skoolwoordeboek die nodige kundigheid deur aktiewe konsultasie met belanghebbendes bekom en dat die woordeboek uiteindelik 'n produk van spanwerk tussen 'n verskeidenheid kundiges eerder as van die werk van 'n enkele leksikograaf sal wees (vgl. Gouws 2006). Wanneer die vraag "Wat hoort in 'n skoolwoordeboek?" gevra word, gaan dit dus nie bloot oor watter data aangebied kan word nie, maar ook oor hóé die data aangebied kan word.

Volgens Bogaards (2003: 28) bestaan daar byna geen navorsing oor woordeboekgebruik deur moedertaalleerders nie. Ten einde meer funksionele skoolwoordeboeke tot stand te bring, sal hierdie beduidende leemte daadwerklik oorkom moet word.

Daar is al gesê dat woordeboeke nie alles vir almal kan wees nie. Wat skoolwoordeboeke betref, kan dit nietemin die leksikograaf se ideaal wees dat die innoverende onderwyser wat oor die regte leestekste vir sy klas en 'n stel optimaal funksionele skoolwoordeboeke beskik, geen ander bron sal benodig om die leerplan suksesvol te dek nie.

Notas

1. Hierdie verwarring kan vererger word deur die huidige gebruik van die term *leerder* in plaas van die vroeëre *leerling* in onderwysgeleedere. Veral in Engels, waar die term *learner* gevolglik

polisemies (en semanties divergent ten opsigte van Afrikaans se *aanleerder* en *leerder*) is, sou die frase "learner's dictionary" in die titel van 'n Suid-Afrikaanse woordeboek na óf 'n aanleerderwoordeboek óf 'n skoolwoordeboek met 'n gepersonifiseerde titel kon verwys.

2. Woordeboekgebruik kan ook as 'n kommunikatiewe handeling beskou word (vgl. Beyer 2006).
3. Vergelykings word wel in die mikrostruktuur van die *HAT Skoolwoordeboek* aangebied, maar nie op konsekwente wyse nie, bv. nie in die artikel van die lemma **arm** nie, maar wel (as "spreekwoord" aangedui) in dié van **vark**.

Bronnelys

Primêre literatuur (woordeboeke)

- Gouws, R.H., M. Stark en L. Gouws.** 2004. *Nuwe woordeboek sonder grense*. Kaapstad: Maskew Miller Longman.
- Gouws, R., L. Potgieter en S. Burgess.** 2010. *Longman Grondslagfasewoordeboek Afrikaans/Engels*. Kaapstad: Maskew Miller Longman.
- Hartmann, R.R.K. en G. James.** 1998. *Dictionary of Lexicography*. Londen/New York: Routledge.
- Louw, P. (Red.).** 2007. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. Kaapstad: Oxford University Press Southern Africa.
- Luther, J. (Red.).** 2009. *HAT Afrikaanse skoolwoordeboek*. Pinelands: Pearson Education South Africa.
- Luther, J.** 2010. *HAT Aktiwiteite en speletjies by die Afrikaanse skoolwoordeboek en Afrikaanse sakwoordeboek*. Pinelands: Pearson Education South Africa.
- Oxford University Press Southern Africa.** 2006. *South African Oxford Secondary School Dictionary*. Kaapstad: Oxford University Press Southern Africa.
- Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns.** 2009. *Afrikaanse woordelys en spelreëls*. Kaapstad: Pharos.

Sekondêre literatuur

- Béjoint, H.** 1989. The Teaching of Dictionary Use: Present State and Future Tasks. Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 208–215. Berlyn/New York: Walter de Gruyter.
- Beyer, H.L.** 2006. 'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelikeid in woordeboeke. *Lexikos* 16: 46–66.
- Bogaards, P.** 2003. Uses and Users of Dictionaries. Van Sterkenburg, P. (Red.). 2003. *A Practical Guide to Lexicography*: 26–33. Amsterdam/Philadelphia: John Benjamins.
- Carstens, A.** 1995. Language Teaching and Dictionary Use: An Overview. *Lexikos* 5: 105–116.
- Combrink, L., N. Faasen, H. Geyser en A. Kloppers.** 1997. *Taalkonteks*. Kenwyn: Juta.
- De Schryver, G., D. Joffe, P. Joffe en S. Hillewaert.** 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67–83.

- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, R.H.** 2006. Leerdergedrewe leksikografie. *Tydskrif vir Taalonderrig* 40(1): 72-84.
- Gouws, R.H.** 2009. Geïntegreerde woordeboekgebruik van vakwoordeboeke vir aanleerders. *Lexikos* 19: 72-93.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosh: SUN PReSS.
- Gouws, R.H. en S. Tarp.** 2008. A New Step Forward for South African Learners' Lexicography: *The Oxford Afrikaans-Engels/English-Afrikaans Skoolwoordeboek/School Dictionary*. *Tydskrif vir Taalonderrig* 42(1): 65-79.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Essex: Pearson Education.
- Hausmann, F.J.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.
- Hiles, L.** 2010. *Examples in South African School Dictionaries: From Theory to Practice*. Ongepubliseerde M.Phil.-thesis. Stellenbosch: Universiteit Stellenbosch.
- Jackson, H.** 2002. *Lexicography. An Introduction*. London/New York: Routledge.
- Kernerman, I.J.** 2007. What's So Good or Bad about Advanced EFL Dictionaries? Gottlieb, H. en J.E. Mogensen (Reds.). 2007. *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004*: 139-145. Amsterdam/Philadelphia: John Benjamins.
- Lombard, F.J.** 1990. 'n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- Louw, P.A.** 2004. *Criteria for a Multifunctional, Monolingual Dictionary in Junior Secondary Education*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- Meij, K., T. Kühn, en R. Snyman.** 1985. *Vakdidaktiek: Afrikaans Moedertaal in die Sekondêre Skool*. Pretoria: De Jager-HAUM.
- Ministry of Basic Education and Culture.** 1998. *Pilot Curriculum Guide for Formal Senior Secondary Education*. Okahandja: National Institute for Educational Development.
- Ministry of Basic Education and Culture.** 2000. *Upper Primary Phase First Language Syllabus: Afrikaans. Grades 5-7*. Okahandja: National Institute for Educational Development.
- Ministry of Basic Education, Sport and Culture.** 1996. *Pilot Curriculum Guide for Formal Basic Education*. Okahandja: National Institute for Educational Development.
- Ministry of Basic Education, Sport and Culture.** 2004. *The Curriculum for the Lower Primary Phase (Grades 1-4)*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2005. *Lower Primary Phase Syllabus Guide. English First Language Grades 1-4*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2006. *Namibia Senior Secondary Certificate Ordinary Level (NSSCO) Afrikaans First Language Syllabus. Grades 11-12*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2006a. *Namibia Senior Secondary Certificate Higher Level (NSSCH) Afrikaans First Language Syllabus. Grades 11-12*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2007. *Junior Secondary Phase Afrikaans First Language Syllabus. Grades 8-10*. Okahandja: National Institute for Educational Development.

- Ministry of Education.** 2008. *National Subject Policy Guide: English First Language Grades 5–12*. Okahandja: Ministry of Education.
- Ministry of Education and Culture.** 1993. *The Language Policy for Schools. 1992–1996 and Beyond*. Windhoek: Ministry of Education and Culture.
- NIED.** 2003. *Learner-Centred Education in the Namibian Context: A Conceptual Framework*. Okahandja: National Institute for Educational Development.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.
- Tarp, S. en R.H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Van der Merwe, M.F.** 2009. Die opvoedkundige waarde van woordeboeke: Voorstelle vir woordeboekonderrig in Suid-Afrika. *Lexikos* 19: 297-313.

The Utilisation of Outer Texts in the Practical Lexicography of African Languages

Emmanuel Chabata, *African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe (emmanuelchabata@yahoo.com)*

and

Dion Nkomo, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch; and Multilingualism Education Project, Centre for Higher Education Development, University of Cape Town, South Africa (deeouf@yahoo.co.uk)*

Abstract: In this article, an analysis of the utilisation of outer texts in recently compiled dictionaries of African languages is presented. The analysis is undertaken in the context of an upsurge of the compilation of dictionaries in the African languages. It is undertaken with a view to do a qualitative evaluation of the many new dictionaries of African languages that have come on the market in recent years. The point of departure is that prior to the recent lexicographic developments, the then available dictionaries were compiled in the context of the limited role the African languages played. It has been found that most of those dictionaries are limited in scope, perspective and function and hence less effective now that the languages are being assigned a greater social role. The evaluation of outer texts in modern dictionaries is therefore a way of measuring the extent to which lexicographic practice in the African languages is applying theoretical developments to produce better dictionaries.

Keywords: LEXICOGRAPHY, LEXICOGRAPHIC PRACTICE, METALEXICOGRAPHY, DICTIONARIES, AFRICAN LANGUAGES, FRAME STRUCTURE, OUTER TEXTS, FRONT MATTER, BACK MATTER, MIDDLE MATTER, LEXICOGRAPHIC DATA

Opsomming: Die aanwending van buitetekste in die praktiese leksikografie van die Afrikatale. In hierdie artikel word 'n ontleding van die aanwending van buitetekste in onlangs saamgestelde woordeboeke van die Afrikatale aangebied. Die ontleding word gedoen in die konteks van 'n oplewing in die samestelling van woordeboeke in die Afrikatale. Dit word onderneem met die oog daarop om 'n kwalitatiewe beoordeling van die baie nuwe woordeboeke van die Afrikatale te doen wat in die laaste jare op die mark gekom het. Die vertrekpunt is dat, voor die onlangse leksikografiese ontwikkelinge, die toe beskikbare woordeboeke saamgestel is binne die konteks van die beperkte rol wat die Afrikatale gespeel het. Daar is vasgestel dat die meeste van daardie woordeboeke beperk in omvang, dimensie en funksie en daarom minder doeltreffend is nou dat 'n groter maatskaplike rol aan die tale toegeken is. Die beoordeling van buitetekste in moderne woordeboeke is daarom 'n manier om die omvang te bepaal waartoe die leksikografiese praktyk in die Afrikatale teoretiese ontwikkelinge toepas om beter woordeboeke voort te bring.

Sleutelwoorde: LEKSIKOGRAFIE, LEKSIKOGRAFIESE PRAKTYK, METALEKSIKOGRAFIE, WOORDEBOEKE, AFRIKATALE, RAAMSTRUKTUUR, BUITETEKSTE, VOORWERK, AGTERWERK, MIDDELWERK, LEKSIKOGRAFIESE DATA

1. Introduction

Zgusta (1991: 334) described the period of the 1970s as "the Golden Age of lexicography". Indeed, lexicography was experiencing what, at that time, had no precedent in its history. Not only were dictionaries being produced in great plenty, but lexicography was actually establishing itself as a fully-fledged discipline, with theoretical lexicography complementing lexicographic practice. While this was generally evident in other languages of the world, it was not the case with African languages where a few speech communities could only have access to very limited lexicographic products, limited in number, typology and quality. At that time, the available dictionaries were those produced in a pre-theoretical environment, compiled mostly by non-linguists incompetent in the less researched African languages. The dictionaries were also produced in the context of the very limited role played by the African languages in the respective communities in which they are used.

For the African languages, the "Golden Age of lexicography" really began in the 1990s when more and new dictionaries began to be produced with a view to empowering the speech communities. It was also then that African linguists became more interested in theoretical lexicography, starting groundbreaking collaboration with international scholars in the field. The African Languages Lexical (ALLEX) Project, which led to the establishment of the African Languages Research Institute (ALRI), was established in 1992 as a collaborative research venture between the University of Zimbabwe and the Universities of Oslo and Gothenburg to produce, among other tools, mother tongue dictionaries in Shona and Ndebele (cf. Chabata 2007, 2008). As the decade ended, the establishment of National Lexicography Units (NLUs) in South Africa signalled a coordinated approach to lexicographic practice of which one target was to empower the multilingual South African nation, including the indigenous speech communities. Ever since then, lexicographic practice has been continuing in these and other African countries, with metalexigraphy also playing a guiding and supporting role, especially in the form of academic training and contributions in lexicographic journals such as *Lexikos*, always investigating and addressing lexicographic challenges through the application of lexicographic theories.

This article analyses the utilisation of outer texts in some of the modern dictionaries produced for the African languages. It is a qualitative effort to measure the extent to which the theories of frame structure and outer texts are reflected in the dictionaries. As has always been reiterated by scholars, lexicographic theory is necessary for enhancing lexicographic practice and dictionary use (cf. Gouws and Prinsloo 2005: 1; Tarp 2008: 11). This article investigates

how the theoretical insights from dictionary structure studies have been applied to produce functional and user-friendly dictionaries, mainly focusing on the utilisation of outer texts.

2. Lexicographic practice in African languages

In order to fully appreciate the current lexicographic developments in African languages, it is important to bear in mind that several dictionaries were produced in these languages many years ago as part of missionaries' efforts to develop the languages so that they could be used for evangelism. These were complemented by the colonial governments' efforts of standardising and developing the languages, which were meant to educate the Africans so that they could easily be converted to European culture, which was anti-African. For example, Hartmann (1990) indicates that several African languages already had dictionaries at the turn of the last decade of the 20th century. In spite of the recent rise in lexicographic activities, the African languages have even up to now not adequately been provided for. For instance, Gouws (1996: 99) observes:

Although the South African languages have a comprehensive collection of dictionaries, it is not hard to detect imbalances between languages, types of dictionaries and lexicographic systems.

Writing specifically on Zimbabwean dictionaries that have been in existence for quite some time, Chabata (2007: 280) argues that they:

were either limited in scope or lacked the much needed focus towards the development and raising of the status of these (indigenous) languages. ... The dictionaries published were bilingual. Generally speaking, bilingual dictionaries are mainly meant for second-language learners. In the case of the Zimbabwean languages, the dictionaries were compiled for the use of the colonial authorities.

The reasons for this state of affairs are historical. The dictionaries were produced in line with the colonial language policies which allotted very limited roles for indigenous languages, mainly for education and evangelisation. That way, they may be characterised as externally-motivated lexicographic products because they were produced to serve the needs that were mainly determined by external members of the African language speech communities (Nkomo 2008: 10; Gouws 2005: 97; Awak 1990: 10; Busane 1990: 20). A good case in point is Weale's (1903) dictionary whose subtitle states clearly that it was "meant for the use of prospectors and farmers in MaShonaland". At that time, the native speakers of the indigenous languages in Zimbabwe were mainly labourers on the mines and farms owned by colonial settlers. However, another very important reason that needs to be acknowledged regarding the poor quality of most dictionaries compiled during the colonial era is that some metalexigraphic advances applied in modern lexicography were yet to be

made. Most dictionary compilers during the colonial era were missionaries, with some of them being linguists without lexicographic expertise and adequate linguistic competence in the indigenous languages.

With new language policies now set in place in post-colonial African states (even though the new policies have not been fully implemented), more research in the indigenous languages, lexicographic training and metalexigraphic advances to support lexicographic practice, the production of dictionaries in the African languages is currently undertaken in a completely different environment. Not only should the communities be provided with more dictionaries of different types, but also with dictionaries of a higher functional value. The progress that lexicographic practice is now experiencing needs to satisfy both the quantitative evaluation which mainly focuses on the availability of dictionaries and their contents, and the qualitative evaluation which focuses on their functional success in terms of both adequacy and accessibility of lexicographic data. As long as the dictionaries that are produced become superior to their predecessors only in terms of contents and ideology, the intended users will find them difficult to use. Lexicographers would be merely compiling dictionaries which satisfy themselves instead of empowering the users. In the remainder of this article, the utilisation of outer texts in the recently produced African language dictionaries will be examined in order to determine the progress that African lexicography has made in his respect.

3. Metalexigraphic advances in the study of outer texts

A linguistic bias which dominated lexicographic practice and scholarship until the final three decades of the twentieth century continued to hinder progress in the planning and utilisation of outer texts in dictionaries. This is not to suggest that dictionaries always consisted exclusively of word lists. They did contain outer texts which, however, were arbitrarily selected and barely utilised for optimal information transfer to the user. Quite often, the outer texts were regarded as appendages to the dictionaries, and rarely as dictionary components. The acknowledgement that the dictionary is a "big text" or a "text compound" (Gouws 2004: 69) consisting of various kinds of texts and subtexts resulted from conscious efforts of developing lexicographic theory. The first scholar to broaden the scope of lexicography in this regard was Wiegand through a quest to formulate a general theory of lexicography (see, for example, Wiegand 1984). Although he did not affirm the status of lexicography as an independent and fully-fledged discipline, he categorically disposed of the long-held view that it was simply part of linguistics or applied linguistics. Wiegand (1984: 13) argued:

Lexicography is not a branch of so-called applied linguistics. ... Lexicography is at all events, more than the application of linguistic theories and methods or the utilization of linguistic and philological findings.

Lexicography is not a branch of lexicology, and lexicography is by no means theoretically determined by lexicology alone.

In a convincing way, he and other scholars (cf. Bergenholtz and Tarp 2003: 172) demonstrated that in some dictionary types, knowledge from other disciplines would be required in addition to or even exclusive of linguistics. The major difference is that Bergenholtz and Tarp have been uncompromisingly assertive of the status of lexicography as a fully-fledged discipline.

Of immediate relevance to this article is the notion of dictionary structure which Wiegand (1984: 17) identified from the outset as an important element of the general theory of lexicography. It became instructive that besides linguistic knowledge, lexicographers need special training and knowledge in order to produce dictionaries (cf. Gouws 2001). This knowledge would entail an ability to plan how dictionary contents could be distributed and arranged throughout a dictionary as it is now widely accepted that not only the availability of certain lexicographic data matters, but also their accessibility by means of dictionary structure (Svensén 2009: 8-9).

Attention was first given to the arrangement of lemmata as well as the presentation of other data categories by means of lexicographic conventions, indicating a clear furtherance of the long established word list bias (Gouws 2007: 77). These constituted what became known as the word list structure (Wiegand and Hausmann 1989: 333), within which the notions of macrostructure and microstructure had emerged (cf. Wiegand and Hausmann 1989: 328). The acknowledgement that not just the word list constitutes a dictionary introduced what was first called the textual book structure (Wiegand and Hausmann 1989: 330), and later became known as the frame structure (cf. Gouws 2002, 2004, 2007). In addition to the central list which is a compulsory text in any dictionary (Wiegand and Hausmann 1989: 331), the frame structure recognised the incorporation of the outer texts, consisting of the front matter, middle matter and back matter texts as part of the dictionary.

Having been originally conceived within Wiegand's general theory of lexicography, the notion of the frame structure or outer texts was initially applied in the description of certain existing dictionaries, leading certain scholars (e.g. Bergenholtz and Tarp 2003: 193) to argue that Wiegand's theory was contemplative and less useful in improving lexicographic practice. However, more research on dictionary structure and the frame structure in particular, advanced from being descriptive to being transformative. This may be discerned from the following comments of Gouws and Prinsloo (2005: 58):

When *planning a new dictionary* lexicographers have to realise that the functionality of the eventual products exceeds the boundaries of the central text. Both the front matter and the back matter can contain texts, which have a functional role in the presentation of the lexicographic data.

Such a transformative approach may be seen as motivated by the introduction of the notion of the data distribution structure by Bergenholtz, Tarp and Wie-

gand (Gouws 2007: 81). Resultant metalexicographical advances gave way to analytical distinctions between integrated and non-integrated outer texts (Gouws 2004) and later function-adhering and non-function-adhering outer texts (Gouws 2007). Such distinctions may be helpful in the understanding of outer texts in existing dictionaries, as well as in their planning and utilisation in new dictionaries. These distinctions will not be reproduced here. Instead, the remaining sections of this article will focus on its specific contribution, which is to evaluate the extent to which progress has been made in African language lexicography regarding the use of outer texts. In so doing, these distinctions will be employed, where necessary, to support the hypothesis that African lexicographic practice has made significant progress in the utilisation of outer texts to facilitate optimum information transfer to the user.

4. Outer texts in African language dictionaries

An article by Nielsen (2009) reveals that very few metalexicographers understand the concept of outer texts or that a dictionary is a carrier of text types. In an analysis of dictionary reviews in *Lexikos* spanning over eighteen years, Nielsen (2009: 220) observes that:

many reviewers exclude outer texts from their evaluation. Those who do include outer texts treat these texts differently, ranging from mere acknowledgment of existence over a description to a critical evaluation.

This is unfortunate, given that dictionary reviews in academic journals such as *Lexikos* are supposedly written by writers with a better understanding of dictionaries and lexicography in general. However, it may be observed that in some cases, it is actually the dictionary compilers themselves who treat outer texts as mere appendages because of limited efforts to integrate these texts into the functions of the respective dictionaries. Many lexicographers continue to justify the underutilisation of outer texts by the argument, as indicated by Busane (1990: 28), that these texts are less frequently or rarely consulted by users. However, Gouws (2007: 81) reasons that it is normal that outer texts will be consulted less frequently compared to the central texts. For example, the user may need to read the guide to dictionary use once in order to use a specific dictionary for an entire lifetime. Thus lexicographers should continue using the outer texts and also devising more innovations for making them more useful and attractive to users. It is contended that this helps in improving the dictionary skills of the users and the general dictionary culture of language communities.

While there are still many dictionaries in the African languages that give little regard to the value of outer texts, the next subsections will focus on examples of dictionaries that demonstrate the progress already made in the utilisation of outer texts to improve dictionary functionality and user-friendliness. In the dictionaries in question, awareness of user needs and a young cul-

ture of using dictionaries in African languages are furthering this progress on the part of lexicographers. The dictionaries of which the outer texts are considered are classified into general bilingual dictionaries, general monolingual dictionaries and specialised (LSP) dictionaries. Although these typological boundaries are not absolute, it will be observed that the importance of certain outer texts depends on the fact that different dictionary types do not only serve different functions, but that they may also serve similar functions quite differently. This has implications for the nature of the data provided by different dictionaries, their data distribution structures and the utilisation of outer texts.

4.1 General bilingual dictionaries and their outer texts

Although several bilingual dictionaries pairing African languages and major languages such as English and French have been in existence for many years, new bilingual dictionaries are being compiled. Their production is set to continue owing to the limitations of those produced in the past. The need to promote multilingualism in countries such as South Africa, for example, means that bi- and multilingual dictionaries have to be produced to support learners of both/all languages covered by a specific dictionary. For example, the *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polopedi ya Sekolo: Sesotho sa Leboa le Seisimane*, henceforth the OBSDNSE/PPSLS, has been designed to enable "learners to read and write, and to hear and speak effectively and competently in their additional language" (De Schryver et al 2007: x). This subsection focuses on how this dictionary utilises outer texts to complement the main text in the realisation of these dictionary functions.

It could be for this reason that the compilers of the OBSDNSE/PPSLS have extensively employed outer texts as carriers of lexicographic data. Overall, the outer texts add up to 56 numbered pages: 12 pages for the front matter, 28 pages for the middle matter and 16 pages for the back matter. This way, the dictionary displays a "complete extension" whereby the main text is "complemented by both front and back matter texts (Gouws 2004: 69). The table of contents of the OBSDNSE/PPSLS presented below provides an overview of the frame structure of the dictionary.

The front matter provides an overview of the entire dictionary. In addition to the imprint and bibliographical pages, it comprises of three integrated outer texts, namely a table of contents, a text entitled "dictionary features" and an introduction. Firstly, the table of contents is directed at all the dictionary texts and subtexts. By simply looking at it, the user is able to see that besides the Northern Sotho–English A–Z and English–Northern Sotho A–Z sections, which are the main places of lexicographic treatment, there are other important texts namely the study and reference sections. As the main texts, the Northern Sotho–English A–Z and English–Northern Sotho A–Z sections on the contents page are indicated by a bolder font. It can be seen that the Northern Sotho–English A–Z section which ends on page 254 is followed by the study section,

which is entitled "Karolo ya go ithuta" in Sesotho sa Leboa and "Study section" in English. The reference section, "Karolo ya boikgakollo"/"Reference section", follows the English–Northern Sotho A–Z section after page 552.

| Diteng | | Contents | |
|---|------------|---|--|
| Dika tša pukuntšu | iv | Dictionary features | |
| Matseno | viii | Introduction | |
| Sesotho sa Leboa–Seisimane A–Z | 1 | Northern Sotho–English A–Z | |
| Karolo ya go ithuta (e latela letlakala la 254) | s1 | Study section (follows page 254) | |
| Mešongwana ya pukuntšu | s2 s4 | Dictionary activities | |
| E-meile ya semmušo | s7 | A formal email | |
| Lengwalo la semmušo | s8 s9 | A formal letter | |
| Lengwalo la segwera | s10 s11 | An informal (friendly) letter | |
| Melaetša ya elektroniki | s12 s12 | Electronic messages | |
| (Madiri a go se be a tlwaelo a Seisimane) | s13 | Irregular verb forms in English | |
| Popopolelo ye khutswane ya Sesotho sa Leboa | s14 s20 | Northern Sotho mini-grammar | |
| (Kwagatšo ya Sesotho sa Leboa) | s24 | Northern Sotho pronunciation | |
| (Kgatholo ya Sesotho sa Leboa) | s26 | Northern Sotho punctuation | |
| (Mopeleto wa Sesotho sa Leboa) | s26 | Northern Sotho spelling | |
| (Kgatholo ya Seisimane) | s27 | English punctuation | |
| (Mopeleto wa Seisimane) | s28 | English spelling | |
| Seisimane–Sesotho sa Leboa A–Z | 255 | English–Northern Sotho A–Z | |
| Karolo ya boikgakollo (e latela letlakala la 552) | R1 | Reference section (follows page 552) | |
| Diruiwa | R2 | Domestic animals | |
| Diphoofolo tša lešoka | R3 | Wild animals | |
| Diphedi tše nnyane le diphoofolo tša ka lewatleng | R4 | Small creatures and sea animals | |
| Dienywa le merogo | R5 | Fruit and vegetables | |
| Mmele wa motho | R6 | The human body | |
| Dipapadi | R7 | Sport | |
| Mmepe wa Afrika-Borwa | R8 | Map of South Africa | |
| Dipolelo tša Afrika-Borwa le dikgato tša thuto | R9 | South African languages and phases of education | |
| Tshedimošo ya mohola | R10 | Useful information | |
| Dinomoro | R12 | Numbers | |
| Boima le dikelo | R14 | Weights and measurements | |
| Dikarabo | R16 | Answers | |

The OBSDNSE/PPSLs table of contents

Secondly, through the text titled "Dika tsa Pukuntšu"/"Dictionary features", the front matter explains in both Sesotho sa Leboa and English the relevance of the various data types included in the dictionary and the way they are represented.

This is illustrated by means of labelled dictionary articles and entries from real dictionary pages. Finally, the introduction provides, among other information, a guide to dictionary use. Thus overall, the front matter provides the user with a concrete idea of the contents and the structure of the dictionary as well as the access routes of the dictionary. Doing this in both languages is consistent with the bilingualism endeavour of the dictionary.

As noted earlier, the study section is located between the Northern Sotho–English A–Z and the English–Northern Sotho A–Z sections of the dictionary. This location may be occupied by either the back matter of the Northern Sotho–English A–Z section or the front matter of the English–Northern Sotho A–Z section. However, it is neither the back matter nor the front matter of the respective A–Z sections, because it is directed at both. As back matter to the Northern Sotho–English A–Z section, it would be incomplete because the reference section is also directed at the same section. It is thus the middle matter of the entire dictionary. For example, the subtexts "Mesongwana ya Pukuntšu"/"Dictionary activities" are relevant for both the Northern Sotho–English A–Z and the English–Northern Sotho A–Z sections as they seek to develop dictionary skills from native language speakers of both languages.

The middle matter of the OBSDNSE/PPSLS is a partially extended complex outer text in which a table of contents functions as a secondary outer text (cf. Gouws 2004: 70) by exposing and guiding the user to 12 subtexts. Its table of contents is extracted from the main table of contents presented above, falling between the Northern Sotho–English A–Z and the English–Northern Sotho A–Z sections. Reproducing the relevant table of contents is convenient, because the user does not have to refer to the main table of contents once again. This can be viewed as user-friendly.

It may be seen that the provision of the middle matter advances the functions of the OBSDNSE/PPSLS which prevail in both the Northern Sotho–English A–Z and the English–Northern Sotho A–Z sections. Whereas the subtexts given between S13 and S28 are indeed necessary to assist the learners of either language to produce appropriate oral and written texts, the subtexts ranging from S7 to S12 specifically support text production in communicative situations of writing formal emails, formal letters, informal (friendly) letters and electronic messages. Perhaps the only objection which could be raised regarding the middle matter would be the arrangement of its subtexts. Since learners also need information regarding spelling, punctuation and irregular inflection when writing formal emails and letters, the subtexts from S13 to S28 may have been more conveniently presented before those between S7 to S12.

The reference section forms the back matter of the OBSDNSE/PPSLS. Not only is it located after both the Northern Sotho–English A–Z and the English–Northern Sotho A–Z sections, it is also relevant to both sections. Like the middle matter (study section), it is a partially extended complex outer text as it has its own table of contents, also extracted from the main table of contents of the dictionary ("Karolo ya boikgakollo"/"Reference section"). Its subtexts further

advance the text production functions by indicating, for example, how names for the months of the year and days of the week are spelt in the two languages. However, texts such as the map of South Africa, South African languages and other useful information have a cognitive function which does not prevail in the main text. What is important is that the cognitive function is also important here, considering that the dictionary is meant for school learners.

4.2 Monolingual dictionaries and their outer texts

While bilingual lexicography in the indigenous African languages may be traced back to over a century ago, monolingual dictionaries only began to appear in the 1980s. The new dictionary type emerged against a background of the African languages playing a minimum social role owing to a lack of standardised orthographies, vocabulary and terminologies. Thus the monolingual dictionaries have had a standardisation role (cf. Hadebe 2006) and an overall language raising impact (Chimhundu 2005). Although the major contribution that general monolingual dictionaries are making in developing African languages are to be found in their main texts (A–Z sections), some compilers have utilised outer texts in a very complementary and productive manner. A quantitative and qualitative evaluation of the utilisation of outer texts in selected monolingual dictionaries in African languages is one indicator of the progress that African lexicography has made in recent years.

Writing on *Isichazamazwi SesiNdebele*, Hadebe (2006: 113) observed at the time that the dictionary had the largest front matter (47 pages) and back matter (18 pages) among Nguni language dictionaries. Important to note here is that *Isichazamazwi SesiNdebele* was largely shaped by the same socio-historical factors and guided by the same reason for compilation and the same general theoretical framework underlying the production of *Duramazwi reChiShona* and *Duramazwi Guru reChiShona*. These dictionaries and *Isichazamazwi SesiNdebele* share a lot, including the extensive utilisation of outer texts. Accordingly, it is befitting that before commenting on the utilisation of selected outer texts from the three general monolingual dictionaries, a sketch is presented of the frame structure of *Duramazwi reChiShona*, which was actually the first one to appear. When additional outer texts in the later dictionaries are discussed, they are seen as cumulative changes in the utilisation of outer texts which were made as the general compilation framework improved.

| Front matter (i-xxx) in English | Main Text (1-504) | Back matter |
|--|-------------------|-------------|
| Imprint | Aa–Zz | |
| List of editorial staff and participants | | |
| ALRI & Publisher addresses | | |
| Preface | | |
| Acknowledgements | | |

| | | |
|------------------------------------|--|--|
| Request for users' comments | | |
| Introduction | | |
| Outline of Shona alphabet | | |
| Shona noun class system | | |
| Shona verbal extensions | | |
| Lemma selection & Defining formats | | |
| The metalanguage list | | |
| Guide to dictionary users | | |

Table 1: The Frame structure of *Duramazwi ReChiShona*

The first column of Table 1 lists the front matter texts of *Duramazwi reChiShona*, the second column represents the main text while the empty third column indicates the lack of back matter. The dictionary has what Gouws (2004: 69) calls a partially extended frame structure, because all the outer texts complementing the main text are only located in the front matter. What is more vital are the constituent parts of the front matter and the way they complement the main text in the realisation of its functions.

Not all the front matter texts in *Duramazwi reChiShona* complement the main text. Some of those that do not complement the main text include the imprint, which provides biographical information that is useful for library cataloguing and indexing. There are also those texts included for other purposes than the realisation of the dictionary functions, such as the comprehensive list of editors and participants, acknowledgements, the editorial request for users' comments, ALRI's and the publisher's addresses as well as the preface. The editorial request page provides contact details through which dictionary users may direct their queries and convey their suggestions to the editorial staff to ensure that subsequent editions and future dictionaries are made more user-friendly. As these texts do not help in the realisation and execution of the dictionary's functions, they are therefore non-integrated and non-function-adhering outer texts.

The following are the front matter texts that complement the main text of *Duramazwi reChiShona* in fulfilling its functions:

- *Mashoko okuvamba* (Introduction)
- *Mabhii anoshandiswa mubumbiro rokunyora ChiShona* (An outline of the Shona alphabet)
- *Zvidimbu zvezviro nedudziro* (Metalanguage list)
- *Mipanda yamazita* (An outline of the noun class system)
- *Zvivanduriro zvezviro* (An outline of verbal extensions)
- *Nzira dzakashandiswa kusarudza nokutsanangura mazwi ari muduramazwi rino* (Lemma selection and defining formats)
- *Mazano okubatsira vanoshandisa duramazwi rino* (Guide to dictionary use)

The outer texts listed above may be divided into two categories according to the way they complement the main text of the dictionary. There are those texts that are venues of data which, just like the main text, may increase the user's knowledge about specific aspects of the Shona language. They include the outline of the Shona alphabet, the Shona noun class system and the Shona verbal extensions. For example, the outline of the Shona alphabet indicates the letters which constitute the Shona orthography and their pronunciation. This way, the outline provides the dictionary user with information regarding the orthography of standard Shona. This is especially important given the questions often raised regarding the gaps between written and spoken Shona whereby some sounds such as 'l', 'q' and 'th' realised in the spoken forms are not represented in writing. In this regard, the alphabetical outline complements the main text in the standardisation of the language since standard forms of words are preferred to or entered before dialectical forms. The noun class system shows the classes into which Shona nouns are classified. This summary in the front matter provides information that is useful in understanding the numbers forming part of every noun entry in the dictionary. The same applies to the table presenting the different types of verbal extensions, their forms and usage examples. This summary provides a guide to the formation of extended verbs as well as the meaning(s) that each verbal extension adds to the verb base. A closer look at the usage examples shows that the meanings provided are those predictable from the individual meanings of the verb base and the verbal extension. The table thus complements the main text, which does not contain extended verbs with predictable meanings. The assumption behind the exclusion of such verbs from the dictionary is that the summary in the front matter is enough in explaining how extended verbs are formed; hence their inclusion would be a form of redundancy. The outline of the Shona alphabet and other texts which provide useful information about the Shona language in the front matter of *Duramazwi reChiShona* are integrated function-adhering outer texts as they contain data which complements the lexicographic endeavours undertaken in the main text. However, some integrated function-adhering outer texts may serve functions which do not necessarily prevail in the main text.

The other group of front matter texts in *Duramazwi reChiShona* complement the main text by guiding users as quickly as possible to various kinds of dictionary data. They include the metalanguage list and all the other types of texts detailing lemma selection procedures, defining formats and other conventions used in the dictionary. These texts, called integrated non-function-adhering outer texts, facilitate the dictionary consultation procedure so that the user finds the relevant assistance.

The utilisation of outer texts in *Isichazamazwi SesiNdebele* and *Duramazwi Guru reChiShona* which were published five years after *Duramazwi reChiShona* display a remarkable improvement. Not only do these dictionaries utilise fully extended frame structures whereby back matter texts are also included, but more texts complementing the main texts in the realisation of both communi-

cative and cognitive functions are also incorporated. *Isichazamazwi SesiNdebele* contains the following back matter texts:

- *Ukulinganisa A & B* (Weights and Measures)
- *Amagama okuphathwa kwabantu* (Political, legal and administrative terms)
- *Imibala ejayekileyo* (Popular colours)
- *Imibala yezinkomo* (Colours and colour patterns of cattle)
- *Amagama enyama yenkomo* (Names of types of meat)
- *Ukudla kwesintu* (Traditional foods)
- *Impahla yemkulwini* (Traditional kitchenware)
- *Izithelo zeganga* (Wild fruit)
- *Isikhathi ngesiNdebele* (Time referencing in Ndebele)
- *Izifo* (Diseases)
- *Izitho zomzimba* (Human body parts)
- *Imihlobo yezinyoni* (Types of birds)
- *Izibungu* (Insects)
- *Amabizo abafana* (Male names)
- *Amabizo amankazana* (Female names)
- *Amazwe lezizwe ze Afrika* (African countries and their nationalities)

Through these texts, *Isichazamazwi SesiNdebele* provides information which increases the knowledge of the users pertaining to various phenomena. While the dictionary is mainly preoccupied with various aspects of language in the main text, it has gone a step further to include cultural data which secondary school students, who are the primary target users, may find useful in their study of Ndebele and other subjects. These texts illustrate integrated function-adhering outer texts which do not necessarily serve the same functions as those prevailing in the main text. The same situation is found in *Duramazwi Guru reChiShona* which, in addition to more or less the same type of texts, contains the following back matter texts:

- *Zvirungamutauro* (Shona proverbs, idioms and pithy sayings)
- *Madzisho nemitupo nezvidao zvavo kumatunhu anotaurwa ChiShona* (Names, totems and sub-totems of Shona traditional chiefs)
- *Mazwi anoshandiswa kutsanangura zvemaumbirwo emutauro nezveuvaramwe* (Linguistic and literary terms)

Proverbs, idioms and pithy sayings are not usually included in general dictionaries mainly because of the word bias which has dominated lexicographic practice and also because of the uncertainties regarding their best location. In *Duramazwi Guru reChiShona*, they are actually explained in a text which immediately follows the main Aa–Zz section and which is actually regarded as "Chikamu 2" (Section 2) of the main text. It may be difficult to say whether they are

part of the main text or the back matter. However, it needs not be overemphasised that their inclusion gives the dictionary an advantage over many monolingual dictionaries. They present useful linguistic and cultural information to the dictionary users, especially Shona students and researchers. Another advantage to users is that instead of buying a dictionary without these sayings and another compilation specifically dealing with them, acquiring *Duramazwi Guru reChiShona* is handier and cheaper.

The text in *Duramazwi Guru reChiShona* dealing with the names, totems and sub-totems of traditional Shona chiefs and their areas of influence is another outer text which provides cultural knowledge to the users. The text containing the two alphabetical indices of linguistic and literary terms, the first listing Shona terms and their English equivalents while the second performs a reciprocal role, is equally important to students and scholars of Shona linguistics and literature. The inclusion of the latter was well-advised as it came before the compilation of *Duramazwi reDudziramutauo neUvaranomwe*, a Shona dictionary of linguistic and literary terms. It is clear from the commentary that has been offered in this section that monolingual lexicography in the African languages is clearly progressing in line with metalexicographic scholarship regarding outer texts.

Another monolingual dictionary of an African language that displays an innovative and progressive utilisation of outer texts is *Eiwanika ly'Olusoga*, the first monolingual dictionary of the Ugandan language Lusoga. In addition to the front and back matter texts, this dictionary contains a middle matter text. Since the front and back matter texts have been discussed with regard to the Zimbabwean dictionaries, attention is given here to the middle matter texts.

The middle matter in *Eiwanika ly'Olusoga* consists of four main texts which are listed below:

- *Ebyafaayo* (Encyclopaedic information)
- *Ebifunze* (Abbreviations)
- *Enfaanana Y'olulimi* (Grammatical information, referred to as the language portrait in the translated front matter)
- *Okufaanana* (Illustrations)

The importance of the middle matter texts in *Eiwanika ly'Olusoga* needs not be over-emphasised, since the value of similar information has been underlined with respect to other dictionaries. They may also be regarded as integrated function-adhering outer texts. It is the innovative strategy the compiler employed in creating the middle matter out of the respective texts that deserves special attention. Unlike the OBSDNSE/PPSLS which is bilingual, the compiler did not have the advantage of accommodating the middle matter between two A-Z sections. Yet the venues for the middle matter texts were not chosen randomly within the main text but in a very meticulous way. The encyclopaedic information in the text *Ebyafaayo* is "given near the entry of **(e)byafaayo**" (Na-

birye 2009: 636). Abbreviations in the text *Ebifunze* are "listed, translated and then explained ... near the entry **(e)kifunze**" (Nabirye 2009: 636). The text which provides a grammatical outline of Lusoga is introduced near the entry of **gulaama** (grammar) (Nabirye 2009: 636). Finally, a collection of all the pictures in the middle matter are "placed near the entry **(o)ku.faanana** (to resemble/look alike)" (Nabirye 2009: 636). Careful observation will reveal a conceptual link between the contents and/or the title of each middle matter text and the macrostructural entry in whose proximity it is placed. This means that, for example, after looking up the meaning of the word **ebyafaayo**, the user can then access the 'ebyafaayo' about the Basoga people as from the next page. Thus, in addition to the standardisation of the language, the dictionary employs an outer text to impart encyclopaedic, historical and cultural information to the user in a very friendly way. In the case of *Ebyafaayo* and *Enfaanana Y'olulimi*, which have sections and subsections, their tables of contents make the data they contain easily accessible.

Finally, another innovation of the Lusoga dictionary as far as outer texts are concerned is the inclusion of an English translation of the front matter as a back matter text, "Appendix 2". Given that Lusoga is categorised as one of the less documented languages (Nabirye 2009: 178), the translated front matter may be useful to those dictionary users who may have problems understanding any part of the front matter. This is possible with respect to the grammatical terms and the metalanguage which the user may encounter for the first time in the dictionary.

4.3 LSP dictionaries and their outer texts

Another dictionary genre that has emerged in the African languages is the specialised dictionary. The main motivation for the production of LSP dictionaries in the African languages revolves around terminology development, which is will hopefully promote the languages so that they may play a more advanced role in education and other specialised domains. Four such dictionaries, namely *Duramazwi reUrapi neUtano* (Shona dictionary of biomedical terms), *Duramazwi reMimhanzi* (Shona dictionary of music terms), *Isichazamazwi Sezo-Mculo* (Ndebele dictionary of music terms) and *Duramazwi reDudziramutauro neUvaranomwe* (Shona dictionary of linguistic and literary terms), have been produced in Zimbabwe since 2004. The need for this type of dictionary is also recognised in South Africa, with the *Multilingual Mathematics Dictionary* (for Grade R to 6) and *Isichazi-magama seziBalo Sezikolo* (Xhosa mathematics dictionary for schools) being some notable publications. These dictionaries also display a frame structure just like the other types of dictionaries, but, compared to them, more importance is attached to certain subtexts as part of their outer texts. These include lists of subject field experts, bibliographies and illustrations.

All the dictionaries mentioned in the previous paragraph contain, in their front matter, lists of subject field experts who participated in the respective dictionary projects. Although all the general dictionaries discussed in this article contain such text segments, the motivation and net effect are quite different. Since most dictionary compilers are linguists themselves, general dictionaries will be found in which the only consulted experts are computer specialists for the creation and management of the dictionary databases. However, the production of LSP dictionaries requires the input of subject field experts. For example, while the lexicographers at ALRI could independently produce the Shona linguistic and literary terms as they are schooled in linguistics and literature, the consultation of relevant subject experts was a prerequisite in the compilation of *Duramazwi reUrapi neUtano*, *Duramazwi reMimhanzi* and *Isichazamazwi SezoMculo* on subject-specific matters. Accordingly, the respective subject field experts are included in the front matter text segments listing the editorial staff members and the project participants. This is also notable in the *Multilingual Mathematics Dictionary* and *Isichazi-magama seziBalo Sezikolo*. In the former, acknowledgements constitute the longest front matter text because collaborators including mathematics specialists and translators from English into the other ten official languages are listed. In the latter, the imprint mentions the individuals who provided mathematical expertise as editors in the dictionary project. Clearly, the motivation behind the listing of subject specialists in the front matter of LSP dictionaries is to authenticate the dictionaries whose conceptual quality may be questioned or even compromised if lexicographers work on their own.

Another notably important outer text in LSP dictionaries produced at ALRI is the bibliography which lists the texts, dictionaries and other specialised publications used by lexicographers as part of their dictionary basis. This needs to be seen as more than an academic or ethical procedure. The bibliography may provide the user with other possible references for further consultation on a specific topic.

Whereas LSP dictionaries just like general purpose dictionaries contain back matter comprising texts that provide cultural or specialised information such as tables of weights and measures, the back matter texts in the LSP African language dictionaries tend to be subject specific. For example, the conversion tables, multiplication tables, different calendar months (according to Moslem, Indian and Jewish traditions), mathematical formulae and shapes, among other data in the back matter of *Isichazi-magama seziBalo Sezikolo* are all relevant for mathematical and numeral literacy in the context of multicultural South Africa. The same applies to illustrations in the back matter of all LSP dictionaries produced by ALRI. For example, in *Isichazamazwi SezoMculo*, the back matter text "Izingoma ezikhethiweyo" (Selected songs) which also contains Zimbabwe's national anthem, is meant to assist the user in learning how to use the music notation symbols in which these songs are presented. Subject specificity may also be observed regarding the back matter texts in *Duramazwi*

reDudziramutauro neUvaranomwe which are specifically included to promote the development of Shona linguistic and literary terminology and scholarship. On this account, most back matter texts in the LSP dictionaries are integrated function adhering as they advance the functions which prevail in the main texts.

5. Conclusion

Dictionary structures are some of the elements whose understanding and utilisation reflects the progress that lexicography as a discipline has made. On this account and specifically through the planning and utilisation of outer texts, practising lexicographers in the African languages have not only embraced but also contributed to such progress. It has been shown in this article that outer texts are employed to facilitate data distribution and the realisation of lexicographic functions either prevailing in the main text or not. The focus of the article has therefore been limited to quantitative and qualitative evaluation, the former praising the presence of outer texts and the latter considering the data in them. There are other relevant aspects regarding the utilisation of outer texts such as the reinforcement of the network systems between the various outer texts and the main text, as well as the consistency which may increase the efficacy of the entire dictionary. These were not dealt with here, although it suffices to mention that improvement is required in this regard. The dictionaries studied in this article are those which demonstrate the progress made in the utilisation of outer texts. Far too many dictionaries are still lacking in this regard and those referred to here may be considered as examples of good lexicographic practice in the African languages.

References

Dictionaries / Encyclopaedias

- Chimhundu, H. (Ed.). 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. et al. (Eds.). 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Chimhundu, H. and Chabata, E. (Eds.). 2006. *Duramazwi reDudziramutauro neUvaranomwe*. Gweru: Mambo Press.
- De Schryver, G.-M. et al. (Eds.). 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšuu ya Polopedi ya Sekolo: Sesotho sa Leboa le Seisimane*. Cape Town: Oxford.
- Hadebe, S. et al. (Eds.). 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Mheta, G. (Ed.). 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Mpofu, N. et al. (Eds.). 2003. *Duramazwi reUrapi neUtano*. Gweru: Mambo Press.
- Nabirye, M. (Ed.). 2009. *Eiwanika ly'Olusoga*. Kampala: Menha Publishers.
- National Language Services. *Multilingual Mathematics Dictionary: For Grade R to 6*. Pretoria: Department of Arts and Culture.
- Nkomo, D. and N. Moyo (Eds.). 2006. *Isichazamazwi SezoMculo*. Gweru: Mambo Press.

- Wababa, Z., K. Welman and K. Press (Eds.).** 2010. *Isichazi-magama seziBalo Sezikolo*. Cape Town: Cambridge University Press.
- Weale, M.E.** 1903. *Matabele and Malaka Vocabulary: Intended for the Use of Prospectors and Farmers in Mashonaland*. Cape Town: Murray and St. Leger.

Other literature

- Awak, M.K.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Ed.). 1990: 8-18.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Busane, M.** Lexicography in Central Africa: The User-Perspective, with Special Reference to Zaire. Hartmann, R.R.K. (Ed.). 1990: 19-35.
- Chabata, E.** 2007. The African Languages Research Institute: A Milestone in the Development of African Languages. *Lexikos* 17: 278-291.
- Chabata, E.** 2008. Language Development: Progress and Challenges in a Multilingual Zimbabwe. *Southern African Linguistics and Applied Language Studies* 26 (1): 13-26.
- Chimhundu, H.** 2005. *Lexicography and Language Raising: Dictionaries in Zimbabwean Languages*. Public lecture presented as part of the Golden Jubilee Celebrations of the University of Zimbabwe, Harare, 13 April 2005.
- Gouws, R.H.** 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset in Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.
- Gouws, R.H.** 2002. Using a Frame Structure to Accommodate Cultural Data. Emejulu, J.D. (Ed.). 2002. *Éléments de lexicographie gabonaise. Tome II*: 54-69. New York: Jimacs-Hillman.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H.** 2005. Lexicography in Africa. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics*: 95-101. Second Edition. Oxford: Elsevier.
- Gouws, R.H.** 2007. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hadebe, S.** 2006. *The Standardisation of the Ndebele Language through Dictionary-Making*. Harare/Oslo: The ALLEX Project.
- Hartmann, R.R.K. (Ed.).** 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24-25 March 1989*. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Handbücher zur Sprach- und Kommunikationswissenschaft* 5.1-5.3. Berlin/New York: W. de Gruyter.
- Nielsen, S.** 2009. The Evaluation of Outside Matter in Dictionary Reviews. *Lexikos* 19: 207-224.

- Nkomo, D.** 2008. *Towards a Theoretical Model for LSP Lexicography in Ndebele with Special Reference to a Ndebele Dictionary of Linguistic and Literary Terms*. Unpublished M.Phil. Thesis. Stellenbosch: Stellenbosch University.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.
- Wiegand, H.E. and Hausmann, F.J.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-360.
- Zgusta, L.** 1991. Probable Future Developments in Lexicography. Hausmann, F.J. et al. (Eds.). 1989-1991: 3157-3167.

The Role of Dictionaries in the Documentation and Codification of African Languages: The Case of Khoisan*

Andy Chebanne, *Centre for Advanced Studies of African Society (CASAS), Cape Town, South African and Department of African Languages, University of Botswana, Gaborone, Botswana (chebanne@casas.co.za / chebanne@mopipi.ub.bw)*

Abstract: Khoisan speech communities are found in Southern Africa, with Botswana having the greatest ethnic and linguistic diversity. Living in small communities, the Khoisan have always found themselves in a situation of being ethnic and linguistic minorities that are socially marginalized. The Khoisan marginalization is historical, resulting in the Khoisan languages not featuring in education and language development. These languages are underresearched and existing publications are divergent, and not accessible to the speech communities themselves. While some anthropologists, linguists, and a few community organizations have undertaken some development of language resources, as yet these efforts have not had the desired effect. The reasons are that the publications used implement impractical linguistic writing conventions or that no supportive programmes exist to encourage more focused and purposeful research and language promotion through literacy classes. This article will review research and publication materials on the Khoisan languages by arguing that there is a need to document and codify these languages to make them accessible to the speech communities. Because dictionary documentation is suggested as the most viable way to accomplish this, existing lexicographical works will be reviewed to demonstrate their strengths and shortcomings. The article will also indicate why and how dictionaries will be critical in this documentation and codification enterprise. Lemmatization procedures will be suggested to facilitate the capture of the linguistic and the indigenous knowledge system of the Khoisan. In this way the Khoisan languages will be vitalized and research contribute to Khoisan language development.

Keywords: KHOISAN LANGUAGES, UNDERRESEARCHED LANGUAGES, LANGUAGE DOCUMENTATION AND CODIFICATION, DICTIONARY, LEXICOGRAPHICAL WORKS, LEMMATIZATION PROCEDURES, LINGUISTIC AND INDIGENOUS KNOWLEDGE, LANGUAGE PRESERVATION AND PROMOTION, BOTSWANA

Opsomming: Die rol van woordeboeke in die dokumentasie en kodifikasie van Afrikatale: Die geval van Khoisan. Khoisanspraakgemeenskappe word in Suidelike Afrika aangetref met Botswana wat die grootste etniese en linguistiese verskeidenheid het.

* This article was presented as a plenary address at the Fifteenth International Conference of the African Association for Lexicography which took place at the University of Botswana, Gaborone, Botswana, 19–21 July 2010.

Deurdat hulle in klein gemeenskappe lewe, het die Khoisan hulle altyd in 'n situasie bevind van etniese en linguistiese minderhede te wees wat maatskaplik gemarginaliseer is. Die marginalisering van die Khoisan is histories, wat daartoe gelei het dat die Khoisantale nie 'n belangrike plek in opvoeding en taalontwikkeling inneem nie. Hierdie tale is ondernagevors en bestaande publikasies is uiteenlopend, en nie toeganklik vir die spraakgemeenskappe self nie. Terwyl sommige antropoloë, linguïste en 'n paar gemeenskapsorganisasies 'n sekere ontwikkeling van taalhulpmiddels onderneem het, het hierdie pogings tot dusver nog nie die gewenste uitwerking gehad nie. Die redes is dat die publikasies wat gebruik word, onpraktiese linguistiese skryfkonvensies toepas of dat geen ondersteunende programme bestaan om meer gefokuste en doelgerigte navorsing en taalverheffing deur geletterheidsklasse te bevorder nie. Hierdie artikel sal die navorsings- en publikasie materiaal betreffende die Khoisantale beoordeel deur te redeneer dat daar 'n behoefte bestaan om hierdie tale te dokumenteer en te kodifiseer om hulle vir die spraakgemeenskappe toeganklik te maak. Omdat dokumentasie deur woordeboeke aanbeveel word as die haalbaarste manier om dit te bewerkstellig, sal bestaande leksikografiese werke bespreek word om hulle sterktes en gebreke te toon. Die artikel sal ook aanbeveel waarom en hoe woordeboeke krities in hierdie dokumentasie- en kodifikasieonderneming sal wees. Lemmatiseringsprosedures sal aanbeveel word om die vaslegging van die taal- en inheemsekennissels van die Khoisan te bevorder. Op hierdie manier sal die Khoisantale verlewendig word en navorsing bydra tot Khoisantaalontwikkeling.

Slutelwoorde: KHOISANTALE, ONDERNAGEVORSTE TALE, TAALDOKUMENTASIE EN -KODIFIKASIE, WOORDEBOEK, LEKSIKOGRAFIESE WERKE, LEMMATISERINGS-PROSEDURES, TAALKUNDIGE EN INHEEMSE KENNIS, TAALINSTANDHOUDING EN -VERHEFFING, BOTSWANA

1. Introduction: Who and why the Khoisan?

The linguistic and ethnographic distinction between Khoe and San is not new and dates back to the early 20th century (Schultze 1928; Barnard 1992; Vossen 1986). The word *Khoe* (in historical texts *Khoi*), meaning "person" for most Khoekhoe (in historical texts, Khoikhoi) speakers has become a generally accepted term for the people and their languages, while the word *San*, which is a Khoekhoe word for "gatherers", comes from *Saon* (with a Nama gender common plural) (Barnard 1992). In recent years, however, the Working Group of Indigenous Minorities in Southern Africa (WIMSA) (Le Roux 1999; WIMSA 2000) has championed efforts to provide Khoe and San languages with a common social reference, the term *San* now being used to transcend ethnic and linguistic distinctions. This may be problematic according to sociohistorical analysis of the label (cf. Chebanne 2003). However, since the name San has the marginalized communities' approval, it could be used advisedly to obviate the historical prejudices associated with it or with some of the communities. The term *Khoisan* is now generally accepted as a common reference for the combined Khoe (those speaking Khoekhoe languages) and San (those speaking non-Khoekhoe languages). Whenever Khoe and San languages are used as separate labels they invariably refer to linguistic entities.

The Khoisan as ethnolinguistic group has an internal diversity that has often led to debates on how exactly the classification should be presented (cf. Güldemann and Vossen 2000). The generally accepted classification of the language families for Khoisan are as follows (adapted from Güldemann and Vossen 2000; Vossen and Keuthmann 1986; Vossen 1988, 1998; Chebanne 2003).

- Proto-Khoe-San**
- 1 **Hadza** (language family, Tanzania)
- 2 **Khoe-San**
- 2.1 **Sandawe** (language family, Tanzania)
- 2.2 **Khoe-San** (Southern Africa)
- 2.2.1 **Khoe-Kwadi** (language family, Botswana and Namibia)
- 2.2.1.1 **Khoekhoe:** Nama and Damara (Khoekhoegowab); !Ora; Hai||om
- 2.2.1.2 **Kalahari Central:** ||Gana; |Gui; Naro
- 2.2.1.3 **Kalahari North:** Khwedam (Kxoe, ||Ani, Buga, |Ganda)
- 2.2.1.4 **Kalahari East:** Kua, Tshoa (Tshwaa), Shua, Tciretcire, Ganadi
- 2.2.2 **San** (language family, Botswana, Namibia and South Africa)
- 2.2.2.1 **Southern San:** Taa and Tuu: !Xóǀ (!Aa): †Ama Oam (†Ama Pfam); !Gwan Oam (!Gwaa Pfam); !Kui (|Xam, ||Xegui, and †Komani)
- 2.2.2.2 **Northern San:** !Xü (!Xung (!Kung)), Ju|'hoan (Ju|'hoasi), †Kx'au||'ein; †Hoa.

The above classification shows the genetic diversity of these ancient languages. Most of the communities who speak these languages are found in Botswana. Also, most of these communities are small and therefore prone to undergoing language shift, dearth and death (cf. Batibo 1997). It is therefore vitally important that these languages should not be left to die before they have even been documented. While the diversity that these speech communities represent within a geographical space gives rise to challenges for linguistic description and anthropological analyses, research should be focused on documentation and codification, and presented in forms readily utilizable by the speech communities themselves.

2. The state of Khoisan anthropological and linguistic research

In Southern Africa, the speech communities of the Khoe and the San are generally invisible in social domains (such as language planning and education). As such they can be considered marginalized and underresearched. However, most of the modern understanding of the Khoe and San is credited to Schapera (1930) and Westphal (1956) who described the ethnography and the ethnolinguistic typology of the Khoisan. The classification that is retained here is mainly adapted from Güldemann and Vossen (2000), whose research has contributed a greater linguistic understanding of Khoisan and its internal linguistic

typology. Thus, the largest part of the relevant research on Khoisan which provides a fair understanding of these language communities has been essentially generated by anthropologists and linguists (cf. Chebanne 2008; Willet et al. 2002). Historical and linguistic works that have resulted in the refined classification of the Khoe and San languages by Güldemann and Vossen (2000), have been undertaken by Westphal (1962, 1971) and Köhler (1971, 1981), the most notable works being by Vossen (1997). Köhler's (1971, 1981) hypothesis is that the common structure of word roots, the combination of rare consonants (ejectives) of click and glottal types largely demonstrated that Khoe and San languages had an ancient genetic relationship, and that pursuing such a hypothesis would be scientifically creditable and productive. The research by Güldemann and Vossen (2000) seems to have brought this debate to some consensus.

To address the theme of this article, it is necessary to contextualize the discussion, and to delimit it to some linguistic samples that will allow a quick appreciation of the issues. To achieve this, the following speech communities will be briefly considered as far as documentation and codification are concerned.

2.1 !Xóǀ

The !Xóǀ language has been the subject of linguistic research for many years (Traill 1986). However, the usage of these publications has been limited to linguists and academics, and has therefore not been accessible to native speakers. !Xóǀ (or !Aa) is a Southern non-Khoe language, belonging to the Taa and Tuu branch, whose closest related, but now extinct languages were once spoken in what is at present South Africa. Incontestably, there is some affinity with the now extinct !Kui (|Xam, ||Xegui, and †Komani) (Güldemann and Vossen 2000; Traill 1986). The speakers of !Xóǀ are scattered over a large area stretching from western Botswana to the eastern Namibian border (Andersson and Janson 1997). It has been reported by Hasselbring (2001) that !Xóǀ is also spoken in some parts of the South African province of the Northern Cape. The !Xóǀ live in small groups without much contact with each other, but share a dialect continuum. Research by Traill (1985) suggests that !Xóǀ has two main varieties, namely †Ama Oam (†Ama Pfam) (Western variety), and !Gwaa Oam (!Gwaa Pfam) (Eastern variety). Most !Xóǀ speakers are bilingual, speaking another language with which they are in contact, namely |Gui (north and east), Nama (south west) and Shekgalagadi (Chebanne and Monaka 2005). However, this bilingualism does not include Setswana (Andersson and Janson 1997), which they encounter only in the administration and the education of their children. Linguistically !Xóǀ is the only language in the Khoe and San family with the most elaborate click phenomenon. This extensive click incidence alongside other complex sounds seems to have impeded its development (Chebanne 2003).

2.2 Ju|'hoasi

Ju|'hoasi (Ju|'hoan), also commonly known as Kaukau in Botswana, is spoken mainly in North Western Botswana and Tsumkwe in Namibia (Güldemann and Vossen 2000; Hasselbring 2000). Ju|'hoasi is considered the central variety of Northern Khoisan (!Xū) which extends into southern Angola. It forms a continuum with †Kx'au||ein (the southern branch of Northern Khoisan). However, the Ju|'hoasi spoken in Tsumkwe is not intelligible to those who speak †Kx'au||ein, and should be considered a dialectal variety of Ju|'hoasi in north-west Botswana (Traill 1986; Güldemann and Vossen 2000). The geographical and social conditions form the basis of the differences between the Namibia and Botswana varieties (Chebanne 2008). The language is mostly used at local village and family level, as well as at community meetings (with interpreters). Ju|'hoasi has been studied extensively by Snyman who wrote a grammar guide and dictionary (1970 and 1975 respectively). Another dictionary was compiled by Dickens (1994). Orthography has been in place since 1969 and was updated in 1987 and 1991. There are many native speakers who can now read and write Ju|'hoasi. This is mainly because of the efforts of the Nyae-Nyae Development Foundation as well as sustained linguistic and cultural research, and missionary work (Chebanne 2008).

2.3 Naro

Naro speakers are found mainly in Botswana, and a very few are encountered in eastern Namibia. In Botswana, many Naro speakers now live and work on the Ghanzi Farms. However, they have maintained their way of life. Naro has been studied by many scholars including Barnard (1985), Bleek (1928), and Visser (1994, 2001). The Naro spoken in the west near the Namibian border is said to be slightly different, as it has been substantially influenced by Nama. Historically, Naro belongs together with |Gui and ||Gana, but except for some lexical items here and there, there is not much mutual intelligibility among them. Naro shows differences in grammatical and phonological structure, which suggest that either it has retained the historical forms while others have lost them, or it has acquired them after it separated from |Gui and ||Gana. Work on the Naro Bible translation has greatly contributed to literacy development. A mass of texts have been developed, thus creating a good database for linguistic and lexicographical work.

2.4 Nama and Damara (Khoekhoegowab)

The languages Nama and Damara are listed under Khoekhoegowab which can be extended to include !Ora. These are mainly Namibian languages. However, in Botswana, they are all identified by their speakers as Nama. In Botswana,

Nama varieties are referred to with Setswana names such as Sekgothu, Sekhikwe and Seqhanakwe (Batibo et al. 2003). Nama has long been codified, its orthography and dictionary dating as far back as 1889. Developed by Kroenlein (Haacke 1999), it has inspired the orthographies of other Khoisan languages. A dictionary has been compiled by Haacke (2003), and there are also grammar guides in Khoekhoegowab (Batibo et al. 2003). Books and texts for primary schools are available, because, according to the Namibian language policy, the first three years of education (Grades 1–3) should be in the mother tongue, this having boosted the resilience of Khoekhoegowab.

2.5 |Gui and ||Gana

The |Gui and the ||Gana languages are mutually intelligible, although the speakers see themselves as different (cf. Nakagawa 2006). The |Gui consider ||Gana as "black Khoe", because of the rather darkish skin of the speakers. Though they have adopted Shekgalagari cultural practices, the people are still relatively positive towards their language (Barnard 1982). In spite of having lived alongside the !Xóǀ speakers, there are only lexical borrowings, but no language switch from |Gui to !Xóǀ. However, the speakers use Naro and Shekgalagari as *lingua franca* (Chebanne 2003) which is a threat to the currency of their own language. This seems to work against the maintenance of the |Gui language as only elderly people, 30 years of age and above, still use it effectively among themselves. Like most of these languages, there is no orthography on which to base any literacy material. Only a few descriptive and socio-historical studies, particularly by Tanaka (1978) are available.

2.6 Khwedam (Kxoe)

The Khwedam (Kxoe) is situated in Northern Botswana. It is a cluster language made up of the ||Ani, Buga, and |Ganda. In much research by German linguists, they are grouped under the label Kxoe (pronounced Khoe). The north western Khoe mainly occupy the Okavango delta of Botswana, and therefore are sometimes referred to as the flood-plain (or River) Khoe. They are essentially homogenous owing to their geographical location and their shared history in and around the delta, which has influenced their lives for thousands of years. The ||Ani and the Buga are the main Khoe communities in this area. Their languages are mutually intelligible. Substantial descriptive studies by German linguists and anthropologists have been done on the ||Ani under the label Kxoe. The literacy efforts associated with the Penduka Declaration will see ||Ani lumped together with Buga, and so contributing to the Khwedam. This will need careful planning and an awareness of how the languages will benefit Khwedam in terms of literacy development and linguistic studies (WIMSA 2001).

3. From research to documentation and codification: What needs to be done?

In one way or another, languages and cultures have communicated with each other long before the existence of language dictionaries. And indeed, dictionaries cannot replace people's capacity to communicate, especially within the framework of the theory of communication. Other mechanisms such as language policies assist in enhancing this communication between languages and cultures. Dictionaries are excellent pedagogical tools; they are indispensable within the language teaching programme. However, community-based dictionaries also have an informal aspect in their use, and therefore need to be designed for non-expert utilization. While dictionary-using skills must be learnt for dictionaries to appropriately respond to the community needs, users must be made to realise the importance of dictionaries. Policies that facilitate communication between languages need to be in place. If there is no recognition of formal literacy in other languages, there will be generalized illiteracy in the communities speaking indigenous languages. In the UNESCO (1995: 57) report there is the following information which should be noted.

Minorities often find it difficult to participate fully in the activities of societies that favour dominant groups. Sometimes this discrimination is embedded in the legal framework that denies these minorities access to education, employment and political representation. More generally, however, the lack of participation is less a matter of official policy than of everyday practice. The challenge consists in first removing discriminatory barriers and then creating the basis for the empowerment of these minorities.

From this report, it becomes clear that the policies that deny the rights of minorities in areas of culture and education, and that exert domination by linguistic and ethnic hegemonies, create a state of discrimination and denial of access to self-realization. To escape from this vicious circle, the UNESCO World Commission on Culture and Development urged the global community in its 1995 report to adopt a vision for the twenty-first century that would uphold the protection and exercise of cultural rights. This vision of cultural diversity seeks to replace prejudice, discrimination and domination with tolerance, fruitful co-existence and equity. Education must be used to enrich Botswana's cultural diversity. All of the country's languages must be taught to a high standard at primary, secondary and tertiary level. But before this could be achieved all these languages must be documented and codified. These observations point out some practical problems that need urgent attention or consideration:

- in the development of indigenous languages, dictionaries will occupy a central role: they should be the basis for codified languages, uniting different dialects, and promoting translation between languages and cultures;

- in the development of dictionaries, especially in the context of minority languages, there is a need to conjugate efforts at the level of dictionary compilation to render them repositories of linguistic and cultural information and instil a sentiment of common utility;
- in the general development of ethnocultural knowledge, and its understanding or communication, there is a need to utilize existing specialized skills: what are the experiences of other dictionary initiatives inside and outside the country?

Therefore there is also a need to specify what these dictionaries of the indigenous languages want to achieve — should they be monolingual, bilingual, multilingual culturally based, grammatically based dictionaries?

- (a) Languages in the culture systems development will be crucial as instruments for reviving the linguistic and cultural resources of any ethnic community.
- (b) The development of dictionaries for the ethnic languages will ensure a positive way towards democratization and the promotion of social equity and harmony, and indeed mutual respect among all ethnic groups, facilitating cultural appreciation and sharing.
- (c) Culture or ethnoculture documentation should underscore the fundamental need to facilitate the speech community to fully access and utilize its linguistic, cultural and ethnic knowledge.

In the following section, a brief look will be taken at the existing dictionaries of the Khoisan languages and the way they contribute to capacity building in linguistic description and ethnocultural knowledge presentation, making them viable projects of both documentation and codification for language promotion and development.

4. Current dictionaries of the Khoisan languages

The foregoing discussions have shown that some useful research and language promotion activities do exist among some Khoisan communities. However, as has been evident, much of this research is academic and does not have much effect in the Khoisan communities. For languages that are under this risk of dying, better methods of documentation and codification should be put in place. Researchers (cf. Haacke 1998; Kgasla and Tsonope 1988) have also argued that dictionaries afford better documentation and codification because their data could be reutilized for many other language related activities (literacy, grammar, linguistic descriptions). For a dictionary to become a source of documentation, it must be construed and conceptualized as a linguistic and cultural reference book or information database which is readily accessible or

utilizable. Also, and this is important, it should provide all possible information that constitute the linguistic and cultural repertoire of the language it is dealing with. The codification dimension hinges on the aspects of linguistic description and prescription, reducing the dictionary material to a code. This code provides rules for orthography, grammar, words and general language usage. In this regard, a dictionary, whatever its design or purpose, is didactic and its material constitutes describable or analyzable elements, that is, words, which are the informational repertoire of a language. While each of these elements or words is independent, they maintain a linguistic relationship at the grammatical level. It is the lexicographical technique which facilitates the way a dictionary as a source of documentation and codification can be accessed and utilized.

Except for Khoekhoegowab, there is a dearth of Khoisan lexicography in Southern Africa (cf. Haacke and Eiseb 2002). The reasons for this state of affairs are varied, but the main ones are: (a) Khoisan studies do not form a significant part of the research publications used in Botswana; (b) linguistic research tends to be technical and therefore not readily utilizable by many scholars; (c) language policies in Botswana restrict learning and usage of the marginalized languages, and thus limit them to the informal domains; and (d) Khoisan speakers are not literate in their languages. To assess the way linguistic work has contributed in documentation and codification of the Khoisan languages, some dictionary excerpts will illustrate the usefulness and also the shortcomings of some of the lexicographical works. The lemma **child** will be considered for all the languages discussed.

4.1 Traill's !Xóõ–English dictionary

Anthony Traill who devoted his academic life to the linguistic description of !Xóõ, published his !Xóõ–English dictionary in 1984. From its presentation, it becomes clear that it was to serve as a linguistic and ethno-anthropological documentation of the !Xóõ people and their language (Traill 1984: 7).

⊙'āni 2 II/I children, diminutive plural marker of ⊙āa. Also general plural marker of membership in certain human groups (e.g. !Xóǰa-tê ⊙'āni, the !Xóõ people, !āma ⊙'āni western bands; †hūā ⊙'āni, southern bands; !xām-tê ⊙'āni, the first people) and other animated non-humans (e.g. g†xā-bu-tê ⊙'āni, the two mythical young gemsbok; ⊙àje ⊙'āni, young animals). But with most regular plural nouns it signifies diminutive only, so, !qhúū-tê ⊙'āni, is small white people; †áā-tê ⊙'āni, is small Kgalagadi people.

Currently Hiroshi Nakagawa and Andy Chebanne are working on the finalization of the trilingual !Xóõ–English–Setswana dictionary. The grand objective of this trilingual dictionary is to record the !Xóõ language, which took Anthony Traill more than 30 years. For an outsider, !Xóõ is a complex language which requires a keen learner and recorder. Traill did an immense work on this lan-

guage, also significantly contributing to socio-anthropological, and sociolinguistic documentation.

4.2 Dickens' Ju|'hoan–English dictionary

The Ju|'hoan–English dictionary compiled by Patrick Dickens was published posthumously in 1994. As the compiler died in 1989 before completing the dictionary, Anthony Traill who was Dickens' mentor completed it. The key lemma, **child**, used in this discussion for the assessment of the linguistic adequacy of a dictionary is only given as:

Dáàmà, n. child

This entry is too brief to capture the versatility of the word *dáàmà* for it can also be used in semantic derivations. By not including these versatile semantic values of *dáàmà*, the dictionary sacrifices much of the linguistic and indigenous knowledge system of Ju|'hoasi.

4.3 Visser's Naro–English dictionary

Hessel Visser's Naro–English dictionary (2001) is grammar and translation focused. However, it contains few other sources of knowledge, especially those based on pure anthropological observation, not prompted by grammar and translation. For the illustrative lemma, **child**, the dictionary presents:

cóá [lɔ́á], 'child'

The Naro dictionary only provides the primary value of the word *cóá*. It does not gloss the semantic derivations the word generates. Here again a lexical work moves on a solely lexical level, not going far in recording the way Naro employs lexical units in the derivation of new lexical terms, and new semantic notions.

4.4 Haacke and Eiseb's Khoekhoegowab dictionary

In 2002, Wilfrid Haacke and Eliphaz Eiseb published their dictionary which represents a significant achievement in the documentation and codification of languages represented, namely Nama and Damara (Khoekhoegowab), and Hai|om. It is an elaborate lexicographical work, providing grammatical information as well as linguistic descriptions. Here follows the treatment of the illustrative lemma, **child**.

Darob, **dawon** [D] {**darob**, **dawob**} *n.* boy; (juvenile) son, s.a |GOAB; **daró-i** {**daró-i**} *n.* child; **daron** {**dranon**} *n.* children; **daros** {**daros**} *n.* girl; (juvenile) daughter; **darorob** {**darorob**} *n.* baby boy; small boy (in size); **daroro-i** {**daroro-i**} *n.* baby; small child; **daroros** {**daroros**} *n.* baby girl; small girl.

4.5 Nakagawa's word lists of |Gui and ||Gana

Hiroshi Nakagawa has made extensive linguistic descriptions of the |Gui and ||Gana language cluster. In 2006, he completed his thesis on aspects of the phonetic and phonological structure of the |Gui language, which have hugely contributed to the understanding of the click sound system in |Gui in particular and Khoisan in general. Although he has compiled some lexicographical works (a |Gui–Japanese word list and a |Gui–Japanese–English word list), he has not yet published a dictionary.

4.6 Kilian-Hatz's Khwe dictionary

This dictionary of Christa Kilian-Hatz, dealing with Khwedam (||Ani and Buga), was published in 2003. It came complete with a practical orthography, notes on grammar, and appendices and it also contains a supplement of Khwe place names in West Caprivi compiled by Matthias Brenzinger. It is an example of a lexicographical work that contributes to the documentation and codification of the Khoisan languages. For the illustrative lemma, **child**, this dictionary presents the following:

ǀóǃ *n* (ǀóǃ ǀóǃ as a second part of a n–n compound; var. ǀóé; q.v.) 1. child (boy, girl), embryo. 2. Son, daughter. 3. Niece, nephew [i.e., child of brother]; cf. -ǀóǃ.
ǀóǃ-dámàcí *n* afterbirth, placenta (*lit.* younger brother/sister of the child);
ǀóǃ-khòè dǀvèè *hon.* 2. *m/f.sg* [from older to younger men/women]; cf. ǀúu-can
ǀóǃ-ngu *n.* boys' hut
ǀóǃ-ò *n.* childhood

The strength of this Khwe dictionary lies in its judicious application of lexicographical techniques by adding to the definitions descriptions and translations. Yet illustrations are kept within limits, not yielding to unnecessary semantic digressions. Indeed, this dictionary has made and will make an important contribution to the efforts for language development in the Khwe community by adopting and justifying the practical orthography which has already been accepted by the speakers (cf. WIMSA 2001). There is therefore no need for a conversion from phonetic to practical as the symbols used constitute the conventional orthography.

Unfortunately, nothing much exists in the other Khoisan languages of Botswana. There are descriptive linguistic and sociolinguistic works on some of these languages (Shua, Tciretcire, Tshoa (Tshwaa) and Kua) but there has as yet been no serious lexicographical undertaking to contribute to documentation and codification. This situation should be viewed with serious concern as most of these languages are spoken in areas where language shift and language attrition are a daily experience.

5. Khoisan dictionary and lexicography work: The way forward

The foregoing observations on the lexicographical content of the current dictionaries present issues of critical concern, which are: (a) some of these dictionaries have been designed as research documents for the sole utilization by linguists; (b) generally these dictionaries have not been planned to contribute to the documentation of the indigenous knowledge of the Khoisan communities; (c) lemmas in these dictionaries are spelt according to the IPA which may not be practical for ordinary writing of the languages; and (d) most of these dictionaries are not accompanied by appendices (grammar, orthography) that facilitate using them as resources for language activities. Because these dictionaries are not monolingual, it means that speakers of these languages should be literate in English and be linguists themselves to access them. This is a critical situation. Therefore the next consideration is: how should Khoisan lexicography proceed to remedy these problems? The other crucial question is: what type (monolingual, bilingual, trilingual, etc.) would facilitate the utilization of Khoisan dictionaries?

For Khoisan communities in Botswana, monolingual dictionaries which lemmatize words in one language may be problematic as these languages are not taught in schools. Bilingual dictionaries which use two languages to lemmatize words are viable but it would require that one of the languages should be accessible to the speakers. Multilingual dictionaries which treat more languages can be a benefit if Setswana is included as it provides a common literacy base for many people in Botswana. Encyclopaedic dictionaries which focus not on just mere words but on concepts in the domain of the function and reference of the word would be ideal for Khoisan. However, the work that would go into an encyclopaedic compilation is huge and nor the Khoisan communities nor dedicated linguistics have resources to realize such a lexicographic undertaking.

In the African languages, lexicographic work has generally not benefited from corpus planning and therefore the quality of the dictionaries varies a great deal (cf. Otlogetswe 2007). For Khoisan specifically, dictionary work has been very insignificant (cf. Haacke 1998). The fact that Khoisan languages exist as marginalized minorities, means that within the countries where they are spoken, they are considered insignificant. Consequently, they are underresearched, and do not feature in any literacy programme. In the main, dictionary material development has not directly resulted from lexicographical work undertaken by researchers, but as a secondary consideration to linguistic data analysis by linguistic researchers, anthropologists and missionaries. These are therefore better qualified as research notes or as glossary materials, especially of a bilingual nature. Later these limitations will be critically considered, and the fact that foreign languages, definitely causing cultural and linguistic obstacles, were and are still used in this kind of work.

The foregoing observations show decisively that if these languages are to be preserved, as they are indeed suffering from the serious assailing effects of language shift, documentation through dictionaries should target language as a vehicle of the cultural and linguistic experience of the speakers. The way out of this situation is that dictionary work should be planned to facilitate functional documentation and codification. The question of a practical orthography is important if this documentation and codification should directly benefit Khoisan speech communities. The current inconsistent and eccentric writing conventions used by some researchers and community language promotion groups compound documentation and codification. The Centre for Advanced Studies of African Society (CASAS) published an orthography for Khoe and San languages developed by Namaseb et al. (2008) with a view to harmonize these divergent writing practices in the documentation and codification of Khoisan languages. However, this work cannot immediately eliminate this orthographic problem, but it should be a reminder to all Khoisan linguistic and ethnocultural promoters to be aware of the need to codify these languages for the ease of cross-Khoisan studies.

Generally, Khoisan languages suffer from poor corpora and the use of non-conventional lexicographical presentations. These linguistic and lexicographical structure problems do not facilitate the exploitation of the existing materials for language development and literacy. The following paragraphs will elaborate on what needs to be done by linguistic researchers and language associations.

One interesting development that facilitates capturing this sort of linguistic and ethnocultural information is the software of the Summer Institute of Linguistics FieldWorks. It is designed to record and generate data in dictionary format. Fieldworks proposes language projects which begin with a careful study of the language and culture of an ethnolinguistic group by:

- data capturing of linguistic and cultural realities observed in practice,
- data analysis and ascertaining linguistic and cultural rules and pre-concepts, and
- generating reports, research papers, dictionaries and write-ups of linguistic and cultural phenomena.

This software helps to manage linguistic and cultural data from the collection point up to the time the linguists wish to publish it (<http://www.sil.org/computing/ddp/DDP>). Chebanne (2007: 22) suggested a keyfile design for a toolbox which would integrate semantic and inter-Khoisan glosses to enhance the study of cognates and the analyses of linguistic and cultural material. Also it is important that the use of foreign languages to define Khoisan entries be adopted with circumspection as the speakers themselves are generally illiterate in these foreign languages. However, this observation also points to the very difficulty of working on these languages and the kind of materials that will

facilitate cross-Khoisan studies. The following keyfile design for Toolbox can be used to input multilingual entries. At the level of formatting there is a choice to generate monolingual, bilingual or trilingual dictionaries.

```
\lx  
\ps  
\sem 1  
\sem 2  
\sem 3  
\ge  
\gs  
\etc
```

(where \lx = lemma in a Khoisan language; \ps = part of speech; \sem 1 ... \sem 3 = to semantic domains; \ge = to gloss English; and \gs = to gloss Setswana).

Note: The back slash (\) is the Toolbox code.

This database design would ensure a fairly informative lemmatization of Khoisan language entries and create access to other Khoisan languages and facilitate inter-Khoisan comparative linguistic and cultural studies. Corpus design is not just a lexicographical, but also a linguistic and cultural imperative. For the Khoisan whose ecology, culture, and languages are endangered, there is a need for lexicographical work that will be holistic in data gathering techniques. While grammar and usage are important for any language, it is the other aspects of life that make the human experience complete. The best approach will be to work, in the first instance, on a bilingual dictionary frame, favouring inter-Khoisan bilingualism: !Xóõ–|Gui, Naro–Nama, etc. The second stage may then consider a multilingual word-list for ease of reference. This way, it will be easier to check lexical adequacies, and lemmatization strategies. But this approach, in contrast to a monolingual dictionary, faces serious semanticization problems when cultural vocabulary items such as proverbial expressions are encountered.

6. Conclusion

Languages such as those belonging to the Khoisan ethnic communities possess an important indigenous knowledge system. The Khoisan communities' knowledge of the environment is complex, their adaptation to the rough desert country require a people with special practical knowledge and skills. This impressive knowledge is found in their understanding of botany, zoology, geography, medicine, and indeed, life-saving skills. This is evident in their interaction with their environment and their ingenuity in times of hardship. The documentation of all these aspects is critical in the survival of these languages and cultures. Dictionaries are favourable for this kind of undertaking because they can holistically document language and culture and thus form a

codified base. Generally, in a region such as Southern Africa, and particularly in a country such as Botswana, codifying these languages will form part of the historical and cultural patrimony of regional countries. Archaeologists are interested in the rock art, artists in their cultural artefacts, but it behoves linguists (lexicographers, grammarians, philologists) to effectively vitalize these languages, and use all methods that could contribute to the recording, documentation and preservation of Khoisan.

References

- Andersson, L.G. and T. Janson. 1997. *Languages of Botswana*. Gaborone: Longman Botswana.
- Barnard, A. 1982. Universal Kin Categorization in Four Bushman Societies. *L'Uomo* 5(2): 219-237.
- Barnard, A. 1985. *A Nharo Word-list, with Notes on Grammar*. Occasional Publication No. 3. Durban: Department of African Studies, University of Natal.
- Barnard, A. 1992. *Hunters and Herders of Southern Africa: A Comparative Ethnography of the Khoisan Peoples*. Cambridge: Cambridge University Press.
- Batibo, H.M. 1997. The Fate of Minority Languages in Botswana. Smieja, B. and M. Tasch (Eds.). 1997. *Human Contact through Language and Linguistics*: 243-252. Frankfurt: Peter Lang.
- Batibo, H.M., J. Mathangwane and J. Tsonope. 2003. *The Third Language Teaching in Botswana. A Consultancy Report for the Department of Curriculum and Evaluation, Ministry of Education, Botswana*.
- Bleek, D.F. 1928. *The Naron: A Bushman Tribe of the Central Kalahari*. Cambridge: Cambridge University Press.
- Chebanne, A. 2003. The Khoe and San Languages and People of Botswana: The Current Situation. Prah, K.K. (Ed.). 2003. *Silenced Voices*: 59-102. CASAS Book Series 34. Cape Town: CASAS.
- Chebanne, A. 2007. Technology and Language Preservation: Lexicographical Techniques in the Study of Under-researched Languages of Botswana. *Journal of the Linguistics Association of the Southern African Development Community (SADC) Universities, LASU* 2, July 2007: 6-17.
- Chebanne, A. 2008. A Sociolinguistic Perspective of the Indigenous Communities of Botswana. *African Study Monographs* 29(3): 93-118. Kyoto: Centre for South East Asia and Africa Area Studies.
- Chebanne, A. and K.C. Monaka. 2005. San Relocation: Endangerment through Development in Botswana. Crawhall, Nigel and Nicholas Ostler. 2005. *Creating Outsiders, Endangered Languages, Migration and Marginalization*: 101-105. Cape Town: Foundation for Endangered Languages (FEL).
- Chebanne, A. and M. Nthapelelang. 2000. The Socio-linguistic Survey of the Eastern Khoe in the Boteti and Makgadikgadi Pans Areas of Botswana. Batibo, H.M. and B. Smieja (Eds.). 2000. *The Future of Minority Languages in Botswana*: 79-94. Duisberg Papers on Research in Language and Culture 40. Frankfurt/New York: Peter Lang.
- Chebanne, A. and L. Nyati-Ramahobo. 2003. *Language Use and Language Knowledge: Data from the 2001 Population and Housing Census*. Paper presented at the Dissemination Seminar on Population and Housing Census, Gaborone, 8-11 September 2003.
- Dickens, P. 1994. *English-Ju/'hoan, Ju/'hoan-English Dictionary*. Research in Khoisan Studies 8. Cologne: Rüdiger Köppe.

- Haacke, W.H.G.** 1998. A Khoekhoe Dictionary in the Making: Some Lexicographic Considerations. Schladt, Mathias (Ed.). 1998. *Language, Identity and Conceptualization among the Khoisan*: 35-64. Research in Khoisan Studies 15. Cologne: Rüdiger Köppe.
- Haacke, W.H.G.** 1999. *The Tonology of the Khoekhoe (Nama/Damara)*. Research in Khoisan Studies. Cologne: Rüdiger Köppe.
- Haacke, W.H.G. and E. Eiseb.** 2002. *A Khoekhoegowab Dictionary with an English–Khoekhoegowab Index*. Windhoek: Gamsberg Macmillian.
- Hasselbring, S.** 2000. Where are the Khoesans of Botswana? Batibo, H.M. and B. Smieja. 2000. *Botswana: The Future of Minority Languages*: 13-32. Duisberg Papers on Research in Language and Culture 40. Frankfurt/New York: Peter Lang.
- Hasselbring, S.** 2001. *A Sociolinguistic Survey of the Languages of Botswana*. Sociolinguistic Studies of Botswana Languages Series 2. University of Botswana/University of Tromsø Collaborative Basarwa Research Programme. Gaborone: University of Botswana.
- Kgasa, M.L.A. and J. Tsonope.** 1998. *Thanodi ya Setswana*. Gaborone: Longman Botswana.
- Kilian-Hatz, C.** 2003. *Khwe Dictionary. With a Supplement on Khwe Place Names of West Caprivi by Matthias Brenzinger*. Namibian African Studies 7. Cologne: Rüdiger Köppe.
- Köhler, O.R.A.** 1971. Die Khoe-sprachigen Buschmänner der Kalahari; ihre Verbreitung und Gliederung. *Forschungen zur allgemeinen und regionalen Geographie: Festschrift für Kurt Kayser zur Vollendung des 65. Lebensjahres*: 373-411. Kölner geographische Arbeiten. Hrsg. von Freunden und Mitarbeitern. Wiesbaden: Franz Steiner.
- Köhler, O.R.A.** 1981. Les langues khoisan. Perrot, J. (Ed.). 1981. *Les langues dans le monde ancien et moderne. Volume 1: Les langues de l'Afrique subsaharienne — Pidgins et créoles*: 455-615. Paris: Éditions du C.N.R.S.
- Le Roux, W.** 1999. *Torn Apart. San Children as Change Agents in a Process of Acculturation. A Report on the Educational Situation of San Children in Southern Africa*. Shakawe, Botswana: Kuru Development Trust.
- Nakagawa, H.** 2006. *Aspects of the Phonetic and Phonological Structure of the |Gui Language*. Unpublished Ph.D. Thesis. Johannesburg: University of the Witwatersrand.
- Namaseb, L., W. Haacke, L. Davids, B.K. Kure, A. Araes, D. Ortman, N. Fredericks and M.C. Moeti.** 2008. *The Standard Unified Orthography for Khoe and San Languages of Southern Africa*. CASAS Monograph Series 232. Cape Town: CASAS.
- Otlogetswe, T.J.** 2007. *Corpus Design for Setswana Lexicography*. Unpublished Ph.D. Thesis. Pretoria: University of Pretoria.
- Schapera, I.** 1930. *The Khoisan Peoples of South Africa: Bushmen and Hottentots*. London: Routledge & Kegan Paul.
- Schultze, L.** 1928. *Zur Kenntnis des Körpers der Hottentotten und Buschmänner. Zoologische und anthropologische Ergebnisse einer Forschungsreise in westlichen und zentralen Südafrika*. Volume 5, Part 3: 147-227.
- Snyman, J.W.** 1970. *An Introduction to the !Xu Language*. Cape Town: A.A. Balkema.
- Snyman, J.W.** 1975. Some Phonetic and Lexical Aspects of Zu|'hoasi. *Bushman and Hottentot Studies* 2: 61-102.
- Traill, A.** 1984. *!Xóǀ Dictionary*. Quellen zur Khoisan-Forschung 15. Cologne: Rüdiger Köppe.
- Traill, A.** 1985. *Phonetic and Phonological Studies of !Xóǀ Bushman*. Quellen zur Khoisan-Forschung 1. Hamburg: Helmut Buske.

- Traill, A.** 1986. Do the Khoi have a Place in the San? New Data on Khoisan Linguistic Relationship. *Sprache und Geschichte im Afrika* 7: 407-430.
- Visser, H.** 2001. *Naro-English, English-Naro Dictionary*. Gantsi: Naro Language Projects.
- Vossen, R.** 1988. Studying the Linguistic and Ethno-History of the Khoe-Speaking (Central Khoisan) Peoples of Botswana. *Botswana Notes and Records* 16: 16-35.
- Vossen, R.** 1997. *Die Khoe-Sprachen. Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Quellen zur Khoisan-Forschung 12. Cologne: Rüdiger Köppe.
- Vossen, R.** 1998. Historical Classification of Khoe (Central Khoisan) Languages of Southern Africa. *African Studies* 57(1): 93-106.
- Vossen, R. and K. Keuthmann (Eds).** 1986. *Contemporary Studies in Khoisan: In Honour of Oswin Kohler: Part 2*. Quellen zur Khoisan-Forschung 5(2). Hamburg: Helmut Buske.
- Westphal, E.O.J.** 1956. The Non-Bantu Languages of Southern Africa. Supplement to Tucker, A.N. and M.A. Bryan. 1956. *The Non-Bantu Languages of North-eastern Africa*: 158-173. London: Oxford University Press for the International African Institute.
- Westphal, E.O.J.** 1962. On Classifying Bushman and Hottentot Languages. *African Language Studies* 3: 30-48.
- Westphal, E.O.J.** 1971. The Click Languages of Southern and Eastern Africa. Sebeok, Thomas A. (Ed.). 1971. *Current Trends in Linguistics, Volume 7*: 367-420. Linguistics in Sub-Saharan Africa. The Hague: Mouton.
- Willet, S., S. Monageng, S. Saugestad and J. Hermans.** 2002. *The Khoe and San: An Annotated Bibliography, Volume 1*. Gaborone: University of Botswana.
- WIMSA & RDU (UB).** 2000. *Education for Remote Area Dwellers in Botswana; Problems and Perspectives. Report on a Three-day Conference on the Education for Remote Area Dwellers in Botswana*. Windhoek: Research and Development Unit (UB) & Workgroup of Indigenous Minorities in Southern Africa.
- WIMSA.** 2001. *The Penduka Declaration on the Standardization of Ju and Khoe Languages*. Windhoek, April 20-22, 2001.
- World Commission on Culture and Development.** 1995. *Our Creative Diversity*. Paris: UNESCO.

A Bilingualised English Dictionary for Catalan Speakers

Montse Corrius, *Departament de Llengües, Ciències Socials i Jurídiques,
Universitat de Vic, Vic, Spain (montse.corrius@uvic.cat)*

and

Dídac Pujol, *Departament de Traducció i Ciències del Llenguatge, Universitat
Pompeu Fabra, Barcelona, Spain (didac.pujol@upf.edu)*

Abstract: In view of the benefits of both monolingual and bilingual dictionaries, a new type of dictionary which combines the strong points of a monolingual learner's dictionary with those of a bilingual dictionary was designed and introduced into some Catalan primary and secondary schools: the *Easy English Dictionary with a Catalan-English Vocabulary*. This article looks in detail at this new kind of dictionary: its philosophy, the way it functions, its advantages over other kinds of dictionaries and the fact that the monolingual part is specifically aimed at Catalan learners of English. Additionally, the article deals with the pedagogical functions of illustrations as well as the adequacy of the dictionary to the lexical needs and requirements of Catalan learners of English in primary and secondary education.

Keywords: FOREIGN LANGUAGE LEARNING, BILINGUALISED DICTIONARIES, ILLUSTRATIONS, GLOCALISATION, PEDAGOGICAL DEFINITIONS, CATALAN LEARNERS OF ENGLISH, LEXICOGRAPHY

Opsomming: 'n Tweetalig verklarende Engelse woordeboek vir sprekers van Katalaans. Gesien die voordele van sowel eentalige as tweetalige woordeboeke, is 'n nuwe soort woordeboek wat die sterk punte van 'n eentalige aanleerderswoordeboek met dié van 'n tweetalige woordeboek kombineer, ontwerp en ingevoer in sekere Katalaanse primêre en sekondêre skole: die *Easy English Dictionary with a Catalan-English Vocabulary*. Hierdie artikel kyk in besonderhede na dié nuwe soort woordeboek: sy filosofie, die manier waarop dit funksioneer, sy voordele bo ander soorte woordeboeke en die feit dat die eentalige deel spesifiek gerig is op Katalaanse aanleerders van Engels. Daarbenewens behandel die artikel die opvoedkundige funksies van illustrasies, asook die leksikale behoeftes en vereistes van Katalaanse aanleerders van Engels in die primêre en sekondêre onderwys.

Sleutelwoorde: VREEMDETAALANLEER, TWEETALIG VERKLARENDE WOORDEBOEKE, ILLUSTRASIES, GLOKALISERING, OPVOEDKUNDIGE DEFINISIES, KATALAANSE AANLEERDERS VAN ENGELS, LEKSIKOGRAFIE

1. Introduction

As has extensively been discussed in the field of foreign language learning, both monolingual and bilingual dictionaries have a series of advantages and disadvantages, which largely condition the choice of one type of dictionary over the other (see Atkins 1985, Fan 2000: 124-125, Laufer and Hadar 1997,

Nakamoto 1995). On the one hand, monolingual dictionaries immerse learners in the second language, but, especially at elementary levels, they cause frustration among users, who often find the definitions difficult to understand. Bilingual dictionaries, on the other hand, provide immediate access to the learner's native language and are essential when someone does not know how to translate an L1 (native language) word in the L2 (foreign language); therefore, bilingual dictionaries generate confidence in the user, even though they do not immerse learners in the L2 as much as monolingual dictionaries.

In order to overcome the disadvantages of both monolingual and bilingual dictionaries and benefit from the advantages of both kinds of dictionaries, a new concept of dictionary, halfway between monolingual and bilingual dictionaries, was devised: the *Easy English Dictionary with a Catalan–English Vocabulary* (EED). This new type of dictionary integrates and interrelates a main L2 dictionary (a monolingual English one) and two bilingual dictionaries (an English–Catalan and a Catalan–English one).

2. The monolingual English dictionary

The monolingual part of the EED has 4 500 entries and 6 000 subentries. Its definitions are meant to be simple and are supported by 1 000 illustrations. At the end of the book, there is also a grammar appendix presenting explanations of the most useful English grammar points and a selection of maps helping users learn the names of the different countries in the world and their corresponding capitals.

The main characteristics of the monolingual English dictionary which make it innovative and easily understood by a Catalan learner of English are: (a) its glocal approach; (b) its clear and pedagogical definitions; and (c) the pedagogical function of its illustrations.

2.1 Glocal approach

Most monolingual learners' dictionaries are aimed at a global market: the same edition is sold in such different countries as Norway, Russia, Spain, Mexico, South Africa and China. This implies that these dictionaries cannot take local linguistic and cultural particularities into account; consequently, they transmit cultural information but encourage cultural homogeneity. Let us take, for example, the definition of 'breakfast' provided by the *Macmillan English Dictionary for Advanced Learners*:

breakfast noun the first meal you have in the morning: *What did you have for breakfast this morning? * I'm never hungry enough for a full breakfast.* **a.** a meal consisting of the things people often have for breakfast, for example eggs or BACON: *Try our all-day breakfast – only £3.50! [...]*

The first subentry states that people often have eggs and bacon for breakfast.

The definition provides useful cultural information, but what should be presented as a local specificity (the definition is clearly British-oriented) is in fact presented as a global characteristic (people in general, and not British people, are said often to have eggs and bacon for breakfast).

The EED, on the other hand, is conceived with a glocal (global + local) approach in mind. 'Glocalization' is a portmanteau word that combines the forms and meanings of 'globalization' and 'localization', and it may be defined as the interaction between the local and the global. Let us consider the example of 'breakfast' as defined in the EED:

breakfast *n*

the meal that you eat in the morning: *Some English people have bacon and eggs for breakfast.*

The EED provides a general definition of 'breakfast' and then, in the example, points towards a culturally specific element, namely the fact that some British people have bacon and eggs for breakfast. Therefore, the example draws attention to a cultural specificity while avoiding cultural homogeneity.

Apart from this, the monolingual part of the EED tends to promote the understanding of definitions and examples through references to Catalan cultural specificities. The example under the entry 'far', for instance, reads: "Miami is very far from Barcelona; Sabadell is very near Terrassa." Indeed, well-known Catalan cities like Sabadell and Terrassa become meaningful examples to the target reader.

Another entry which is also representative of the glocal approach adopted in this dictionary is 'tongue twister':

tongue twister *n*

a group of words difficult to pronounce when you say them quickly: *'Setze jutges d'un jutjat mengen fetge d'un penjat' is a Catalan tongue twister; 'She sells sea shells on the seashore' is an English tongue twister.*

As expected from a good monolingual learner's dictionary, the EED provides users with an example of 'tongue twister': "She sells sea shells on the seashore." But, unlike most dictionaries aimed at an exclusively global market, the EED resorts to the learner's cultural background ("Setze jutges d'un jutjat mengen fetge d'un penjat") as an effective means of conveying the meaning of 'tongue twister'. The EED is in fact conceived to mediate between the English and the Catalan languages and cultures.

2.2 Clear and pedagogical definitions

Except for references to local culture, another aspect which contributes to making the dictionary easier for Catalan speakers is the clear and pedagogical definitions contained in its entries. The conception of the EED, its target users and the trials made in a number of Catalan schools suggested that in the monolingual entries simplicity was to be preferred to lexicographic precision.

A first step towards clear and pedagogical definitions is the use of a restricted vocabulary. The words used in the definitions and examples are basic words that appear defined in the dictionary and that have been chosen, whenever possible and without renouncing naturalness, for their similarity to Catalan words. Below are a couple of examples illustrating this:

cheetah *n*

an animal that is similar to a leopard: *Cheetahs live in Africa and are the fastest animals in the world.*

fortune-teller *n*

a person who predicts the future by looking at cards, at the lines on your hands, etc.

In the definition of 'cheetah', there are some keywords that, because of their similarity to Catalan, help a Catalan speaker to understand the English text. These are: 'animal', 'similar' and 'leopard'. The Catalan equivalents are *animal*, *similar* and *leopard* respectively. In the case of 'fortune-teller', the words similar to Catalan are: 'person', 'predicts' and 'future'. The Catalan equivalents are *persona*, *prediu* and *futur* respectively.

Apart from this, simple syntax is used to facilitate Catalan users understand the meaning of each entry. Thus, complex structures learnt at an intermediate level, such as conditional phrases, passive constructions and relative clauses have been avoided whenever possible.

Furthermore, in order to help learners understand the English text, in most instances there is an example next to the definition or even an illustration with a clear pedagogical purpose. The examples can be found just after the definition or at the foot of the illustration which, as will be seen in the following section, is aimed at clarifying the meaning and usage of the word. In a few instances, just when the definitions seem too complicated, these have been omitted and only an example — with or without an illustration — has been used:

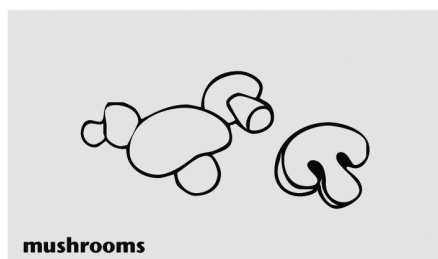
plus *prep*

Two plus two equals four (2+2=4).

mushroom *n* /'mʌʃru:m; 'mʌʃrʊm/

We went to pick mushrooms last Sunday.

667/5979



mushrooms

Figure 1: 'mushroom'

Careful attention has also been paid to the phraseology associated with each word: common lexical combinations, idioms, prepositions, phrasal verbs, etc. Phraseology may be said to be an essential element for the learning of a foreign language, since it shows the most common contexts in which words appear: users, for example, will find the form 'it's raining' under the entry 'rain' because it is a common expression used with 'rain'. A Catalan speaker could make the mistake of saying 'rains' instead of 'it's raining' since in Catalan the subject is not used, so the dictionary helps learners use the right form. A Catalan learner of English should also know that the English say 'to rob a bank' instead of 'to steal a bank', and that they say 'at Christmas' instead of 'in Christmas', so the dictionary reflects this phraseology. When Catalan users look up a word they already know in the EED, they will find additional information: for example, under the entry 'ambulance', which they most probably know because it is very similar to the Catalan form *ambulància*, they will learn that ambulance is used with the verb 'to call' and not with a verb such as 'to warn', as they would say in Catalan. Something similar may happen with the entry 'Internet', the form of which is exactly the same in Catalan. If they look up this word in the EED, they will find additional information they probably do not know. For example, they will find the lexical combinations 'Internet address', 'Internet user', 'to access the Internet', 'to search for something on the Internet' and 'to surf the Internet'. In the case of the last collocation, a Catalan speaker, translating from Catalan, may make the mistake of saying 'to sail for Internet'.

As the EED is particularly aimed at primary and secondary education students, it also includes the essential cross-curricular vocabulary used in primary and secondary schools. Thus, one can find words related to:

- Mathematics: 'addition', 'subtraction', 'centimetre', 'ounce', 'triangle', 'circle', etc.
- Natural sciences: 'eclipse', 'hurricane', 'earthquake', 'rib', 'lung', etc.
- Social sciences: the names of different countries and their capitals, the names of rivers, mountain ranges, etc.
- Technology: materials ('plastic', 'wood', 'cardboard'), tools ('spade', 'shovel', 'rake'), etc.
- New technologies: 'e-mail', 'at' ('@'), 'videoconference', 'cellphone', 'DVD player', etc.
- Language and literature: 'punctuation mark', 'poem', 'play', 'writer', etc.
- Music: 'jazz', 'note', the names of different musical instruments, etc.
- Arts and crafts: 'drawing', 'brush', 'cut-out', 'cardboard', etc.
- Sports: the names of different sports and related expressions ('to score a goal', 'to bounce a ball', etc.).
- Leisure: 'jigsaw puzzle', 'to play marbles', 'disco', etc.

The above words have not only been defined, as would be expected from any dictionary; in some cases, they also provide encyclopedic (geographical, cultural, social, etc.) information. The examples in the entries below illustrate this point:

ocean *n*

a big sea: *There are five oceans in the world: the Atlantic Ocean, the Pacific Ocean, the Indian Ocean, the Arctic Ocean and the Antarctic Ocean.*

war *n*

a fight between two or more countries: *World War II started in 1939 and finished in 1945.*

2.3 Pedagogical uses of illustrations

The illustrations in the EED are in no way meant to be fillers: they have a clear pedagogical purpose. They may have one (or several) of the following functions, assisting in: (i) understanding definitions and focusing on common lexical combinations; (ii) expanding vocabulary; (iii) distinguishing between related words; (iv) grouping together related words; and (v) comprehending onomatopoeias.

2.3.1 Understanding definitions and focusing on common lexical combinations

The most basic function of illustrations is to help users understand the words defined (see Figure 2).

earring *n* /'i:ərɪŋ/
an ornament for your ears 378



Figure 2: 'earring'

Although the definition in Figure 2 includes a basic word like 'ears' and a word identical in form to Catalan ('ornament'), which may facilitate the understanding of 'earring', the illustration reinforces comprehension and is a useful tool to

learn the lexical combination 'to wear earrings'. Lexical combinations are key aspects in language learning, for they reflect the way a given word relates to others. Learning lexical combinations, therefore, is an essential step for the learner to produce natural-sounding, idiomatic language in the target language.

2.3.2 Expanding vocabulary

Illustrations are also used to expand vocabulary, whether it be by contrast (Figure 3), by segmentation (Figure 4) or by contiguity (Figure 5).

smile /smaɪl/

- 1 *v* when we are happy, we smile 5307
- 2 *n* an expression of happiness: *Theresa is very nice. She always has a smile on her face.* 5307



Figure 3: 'smile'

elephant *n* /'elɪfənt/

- a very large animal with big ears and a long trunk: *Elephants come from India and Africa.* 1875

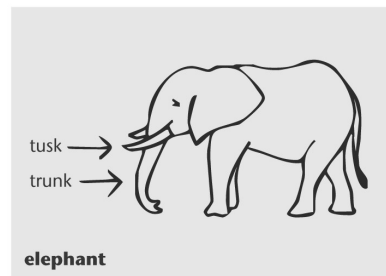


Figure 4: 'elephant'

root *n* /ru:t/

- the part of a plant that is under the ground: *Carrots are roots.* 380/4880

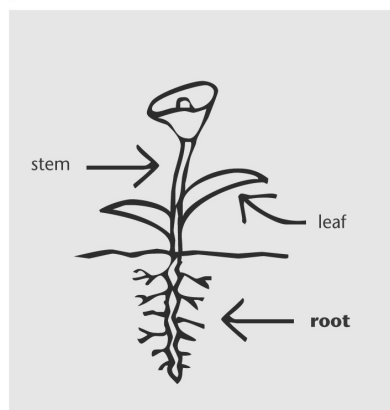


Figure 5: 'root'

2.3.3 Distinguishing between related words

Some illustrations help the learner realise that what in Catalan is expressed by a single word, in English may have several equivalents. The word *escala*, for example, has three different equivalents in English, namely 'stairs', 'escalator', and 'ladder' (see Figure 6).

stairs *n* /steəz/
a group of steps that allow you to go from one floor to another: *to go up the stairs, to go down the stairs, to climb the stairs, etc.* 2081

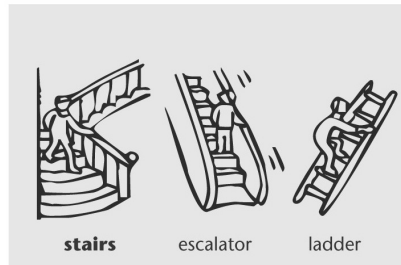


Figure 6: 'stairs'

container *n* /kən'teɪnə(r)/
a box, a bottle, etc. to put things inside
1277

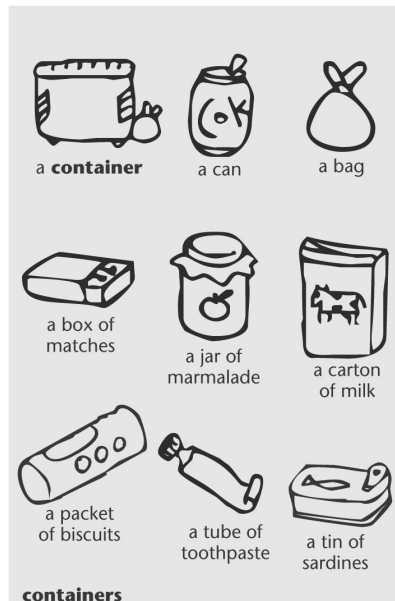


Figure 7: 'container'

2.3.4 Grouping together related words

Some pictures group semantically related words together. Thus, the entry 'insect' includes an illustration with the most common types of insects (a mosquito, an ant, a butterfly and a fly). Likewise, the entry 'container' (Figure 7) illustrates different kinds of containers.

2.3.5 Comprehending onomatopoeias

Figure 8 includes the onomatopoeia for a cock's crow.

cock *n* /kɒk/
a male chicken 2645

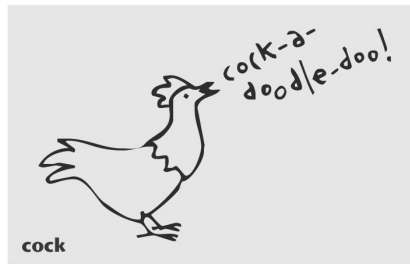


Figure 8: 'cock'

The EED illustrates the most common onomatopoeias, for example 'woof! woof!' (under the entry 'dog'), 'baa! baa!' (under the entry 'sheep'), 'achoo!' (under the entry 'sneeze'), or 'ha! ha!' (under the entry 'laugh').

3. Integration and interrelation of a monolingual and two bilingual dictionaries

The EED is a 'bilingualised' dictionary. This means that translations into the learner's own language also appear in the dictionary. There are two kinds of bilingualised dictionaries: immediate and deferred (see Pujol, Corrius and Masnou 2006). In immediate bilingualised dictionaries the translation comes immediately after the monolingual part, that is to say next to the definition of the word. The following is an example from an immediate bilingualised dictionary, the *Password diccionario didáctico/Password English Dictionary for Speakers of Spanish*:

timetable *n.* list which shows the times of trains/aircraft/classes in school/appointments: *have you the latest train timetable?; there are two English lessons on the timetable today; we have to keep to a strict timetable of appointments.*

□ **horario**

As can be observed, the advantage of immediate bilingualised dictionaries is that users find the translation directly after the definition (in this case, *horario* is the translation of 'timetable'). An objection, though, can be made to this type of dictionary: 'Do the users really read the L2 text?' (Nakamoto 1995: section 5). In other words: do users go straight to the translation, dismissing the L2 part? Two studies point in this direction. The first study is by Laufer and Kimmel (1997: 367): they find that in immediate bilingualised dictionaries the highest percentage (31%) of (immediate) bilingualised dictionary users read only the L1 text, while 27% of users read only the monolingual L2 part. The second study is by Thumb (2004: 91): she discovers that 39% of users read only the L1 translation, thus skipping the monolingual L2 part.

In order to overcome the practice that users skip the monolingual L2 part in immediate bilingualised dictionaries, the deferred bilingualised dictionary was conceived. In this type of dictionary, translations do not immediately follow the monolingual L2 part; rather, they are deferred to the second part of the dictionary. Figure 9 illustrates this.

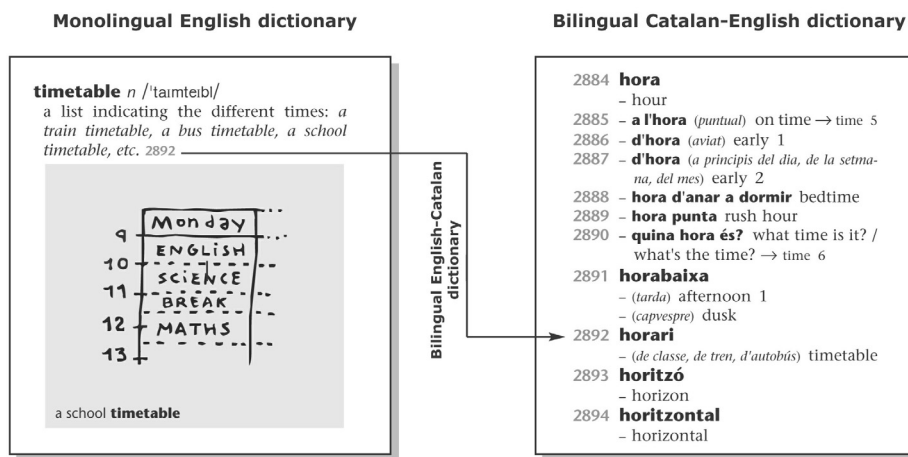


Figure 9: Macrostructure of the EED (1)

In Figure 9, it will be noted that in the monolingual English dictionary at the very end of each entry (or subentry) there is a small number (or a series of numbers) which directs the user towards the Catalan translation of the English word. Thus, in the case where users do not understand the meaning of 'timetable', they can resort to the number 2892 in order to find out, in the second part of the dictionary, its corresponding Catalan translation: *horari*. The interrelation, by means of numbers, between the monolingual English part of the dictionary and the Catalan translations makes up the bilingual English-Catalan dictionary, which in turn acts as a bilingual Catalan-English dictionary.

On occasion, it can be seen that a Catalan lexical combination is translated as an English lexical combination (see Figure 10).

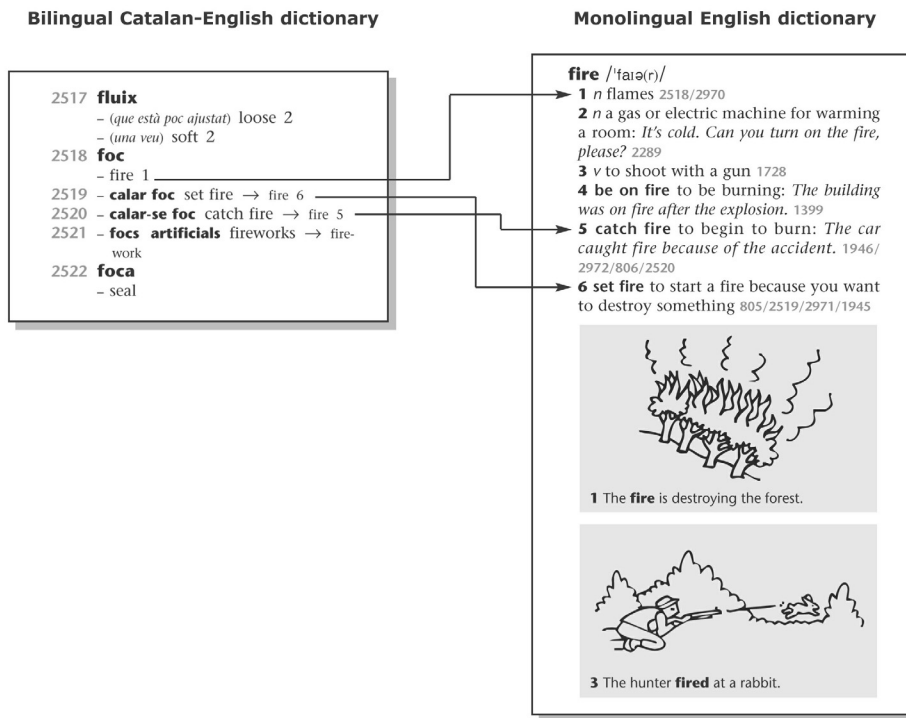


Figure 10: Macrostructure of the EED (2)

In Figure 10, the lexical combination *calar-se foc* is translated as 'catch fire', and the keyword after the arrow ('fire 5') refers users, in case they need further information, especially for oral or written production, to the monolingual English part. The same applies to *calar foc* 'set fire': the keyword refers users to 'fire 6'. Apart from the keywords, another feature that needs pointing out is the inclusion of numbers after certain Catalan→English translations. Thus, the number 1 after 'fire' indicates that this particular meaning of 'fire' corresponds to the first subentry in the monolingual English dictionary.

The EED thus integrates a monolingual L2 dictionary and two bilingual dictionaries: a bilingual L2→L1 (English→Catalan) dictionary and a bilingual L1→L2 (Catalan→English) dictionary. These are not separate units simply brought together in one book, that is: the three dictionaries are not juxtaposed, but are interrelated in such a way that there is circularity. Owing to circularity, the monolingual dictionary (for example 'raise' in Figure 11) takes users to the bilingual (English→Catalan) section (*apujar*), which in turn takes them to the bilingual (Catalan→English) dictionary (*apujar* → 'turn up'; *apujar* → 'increase'),

which further takes them to the monolingual (English) part ('turn 7' and 'increase 1'):

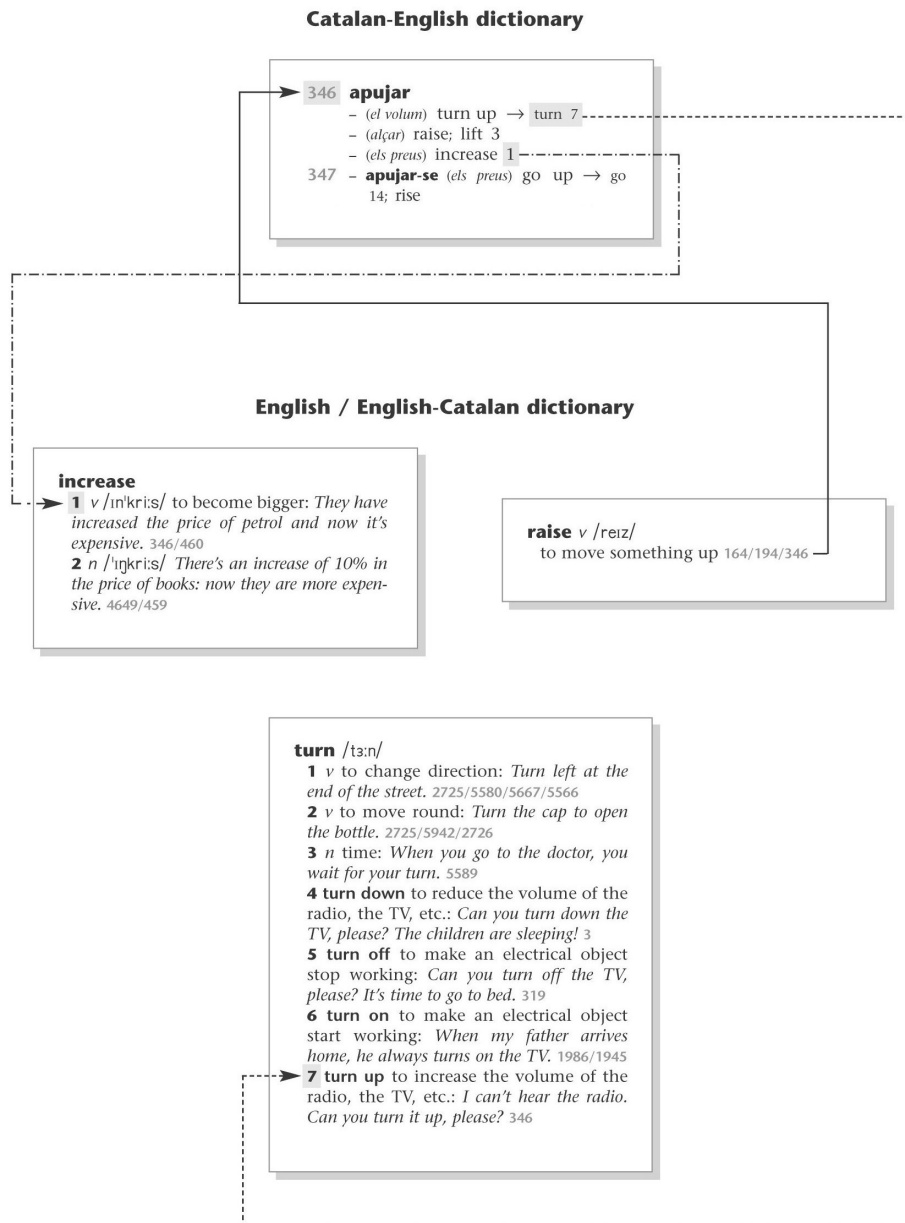


Figure 11: Macrostructure of the EED (3)

4. Conclusion

The EED is a new concept of dictionary that integrates and interrelates three different types of dictionaries: a monolingual L2 one, a bilingual L2→L1 one, and a bilingual L1→L2 one. The dictionary, aimed at Catalan learners of English as a foreign language, consists of three different parts, two of them physically differentiated (the monolingual English dictionary and the bilingual Catalan–English dictionary) and one (the bilingual English–Catalan dictionary) establishing a relation between the other two parts. The main purpose of this type of dictionary is to encourage students to use monolingual dictionaries as much as possible, without denying them the possibility of access to translations. In doing so, learners can benefit from the strengths of both monolingual and bilingual dictionaries.

The monolingual English part takes into account the special needs and requirements of learners of English in primary and secondary education, which is mainly reflected in the clear and pedagogical definitions and the pedagogical uses of illustrations. The monolingual part has also been conceived with a global approach, which helps Catalan speakers understand the definitions in the foreign language and provides both English and Catalan specific cultural information. Additionally, the monolingual English dictionary directs the user towards the bilingual English–Catalan part, which contains the translations. To conclude: the integration and interrelation of three different types of dictionary and the particular conception of both the monolingual and the bilingual parts make the EED a helpful tool for learning English as a foreign language and a model that can be adapted to other languages, for example, English and Afrikaans, or one or other of the African languages.

References

- Atkins, Beryl T. 1985. Monolingual and Bilingual Learners' Dictionaries: A Comparison. Ilson, Robert (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 15-24. Oxford: Pergamon Press/The British Council.
- Easy English Dictionary with a Catalan–English Vocabulary*. 2004. Vic: Eumo Editorial.
- Fan, May Y. 2000. The Dictionary Look-Up Behavior of Hong Kong Students: A Large-Scale Survey. *Educational Journal* 28(1): 123-138.
- Laufer, Batia and Linor Hadar. 1997. Assessing the Effectiveness of Monolingual, Bilingual, and 'Bilingualised' Dictionaries in the Comprehension and Production of New Words. *The Modern Language Journal* 81(2): 189-196.
- Laufer, Batia and Michal Kimmel. 1997. Bilingualised Dictionaries: How Learners Really Use Them. *System* 25(3): 361-369.
- Macmillan English Dictionary for Advanced Learners*. 2002. (Edited by M. Rundell and G. Fox.) London: Macmillan.
- Nakamoto, Kyohei. 1995. Monolingual or Bilingual, That is *Not* the Question: The 'Bilingualised' Dictionary. *Kernerman Dictionary News* 2. <http://kdictionaries.com/newsletter/kdn2-2.html>.

Password diccionario didáctico/Password English Dictionary for Speakers of Spanish. 2002. Sixth Edition. Madrid: SM.

Pujol, Dídac, Corrius Montse and Joan Masnou. 2006. Print Deferred Bilingualised Dictionaries and their Implications for Effective Language Learning: A New Approach to Pedagogical Lexicography. *International Journal of Lexicography* 19(2): 197-215.

Thumb, Jenny. 2004. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary: A Think-aloud Study*. Lexicographica. Series Maior 117. Tübingen: Max Niemeyer Verlag.

Een digitaal compilatiecorpus historisch Nederlands

Evie Coussé, *Vakgroep Nederlandse Taalkunde, Universiteit Gent, Gent, België (evie.cousse@ugent.be)*

Samenvatting: In deze bijdrage wordt een digitaal compilatiecorpus historisch Nederlands voorgesteld dat historische teksten uit verschillende bestaande bronnenverzamelingen samenbrengt tot een methodologisch verantwoorde steekproef die de gehele geschiedenis van het Nederlands overspant. Het compilatiecorpus bestaat uit twee deelcorpora die elk een stuk van de gehele tijdspanne bestrijken en die elk één bepaald type taalgebruik weerspiegelen. Het deelcorpus ambtelijke teksten beoogt het regionaal gekleurde taalgebruik te representeren dat typisch is voor de geschreven taal uit de middeleeuwen en bevat een selectie van lokale ambtelijke teksten uit de drie dialectstreken Vlaanderen, Brabant en Holland voor de tijdspanne van 1250 tot 1800. Het deelcorpus narratieve teksten beoogt dan weer het geschreven taalgebruik in Holland evenwichtig te representeren vanaf het einde van de zestiende eeuw tot vandaag. De evenwichtige samenstelling van het compilatiecorpus maakt het mogelijk om taalveranderingen ononderbroken in de tijd en ruimte te volgen vanaf het vroegste Nederlands tot vandaag.

Sleutelwoorden: CORPUS, CORPUSVERZAMELING, CORPUSMETHODOLOGIE, DIGITALISERING, AMBTELIJKE TEKSTEN, NARRATIEVE TEKSTEN, HISTORISCHE TAALKUNDE, NEDERLANDS, MIDDELNEDERLANDS, NEDERLANDSE DIALECTEN

Abstract: A Digital Compilation Corpus Historical Dutch. In this article, a digital compilation corpus of historical Dutch is presented that brings together historical texts from different source collections in a methodologically motivated sample spanning the whole history of Dutch. The compilation corpus consists of two subcorpora which each covers a part of the complete time span and which each reflects one type of specific language use. The subcorpus of chancellery texts aims to represent the regionally coloured language use typical of the written language of the Middle Ages and contains a selection of local chancellery texts from the three dialect regions Flanders, Brabant and Holland for the time span from 1250 to 1800. The subcorpus of narrative texts aims to give a balanced representation of the written language use in Holland from the end of the sixteenth century to the present day. The balanced composition of the compilation corpus enables the user to follow language changes uninterruptedly in time and space from the earliest Dutch to the present day.

Keywords: CORPUS, CORPUS COLLECTION, CORPUS METHODOLOGY, DIGITALIZING, CHANCELLERY TEXTS, NARRATIVE TEXTS, HISTORICAL LINGUISTICS, DUTCH, MIDDLE DUTCH, DUTCH DIALECTS

1. Inleiding

De taalkundige studie van het historische Nederlands staat of valt met de beschikbaarheid van betrouwbare bronnen die ons toegang geven tot het taalgebruik van vroegere generaties. De voorbije decennia zijn steeds meer historische bronnen digitaal ontsloten, wat hun toegankelijkheid en bruikbaarheid voor taalkundig onderzoek aanzienlijk heeft verhoogd. Het klassieke voorbeeld van een gebruiksvriendelijke collectie met elektronische historische teksten is de cd-rom *Middelnederlands*, die naast het gedigitaliseerde *Corpus Gysseling* (dat alle overgeleverde Middelnederlandse teksten uit de dertiende eeuw verzamelt) ook een selectie van Middelnederlandse rijm- en prozateksten bevat uit latere eeuwen. Daarnaast is de laatste jaren ook het belang van het internet gegroeid bij de digitale ontsluiting van historische bronnen. Ik denk hierbij bijvoorbeeld aan de Digitale Bibliotheek voor de Nederlandse Letteren (www.dbnl.org), die toegang biedt tot een steeds groeiende verzameling klassiekers uit de Nederlandse literatuur in digitaal formaat. Voor een exhaustiever overzicht van andere digitale bronnen verwijs ik naar Coussé (2007).

Hoewel tegenwoordig steeds meer historische bronnen digitaal ter beschikking zijn, is het niet zo dat we een taalkundig fenomeen met wat eenvoudige muisklikken zomaar kunnen onderzoeken vanaf het vroegste Middelnederlands tot vandaag. Om te beginnen is er het praktische probleem dat het bronnenmateriaal zich verspreid over verschillende cd-roms en internetlocaties bevindt. Als we die bronnen allemaal in ons historisch onderzoek van het Nederlands willen betrekken, moet dat materiaal in de eerste plaats dus op één plek gecentraliseerd worden. Daarnaast is er de bijkomende praktische complicatie dat de teksten op de cd-roms en het internet doorgaans niet op dezelfde manier geannoteerd werden (als er al een annotatie beschikbaar is) en/of niet in hetzelfde formaat gedigitaliseerd zijn. Als we met een paar simpele muisklikken dat materiaal willen doorzoeken, moet er dus gestreefd worden naar een uniforme annotatie en codering van de digitale gegevens. Ten slotte is er nog het methodologische probleem dat het voor ernstig taalhistorisch onderzoek doorgaans niet volstaat om zomaar wat historische teksten samen te brengen. Voor betrouwbare resultaten is het van belang dat er een evenwichtig corpus samengesteld wordt, waarbij de invloed van buitentalige variatie (bv. uiteenlopende herkomst of datering van de teksten) uit het corpus is geweerd of op een gecontroleerde manier in de tekstverzameling is geïntegreerd. We moeten dus met andere woorden een betrouwbare steekproef samenstellen van historische teksten die een welbepaald soort historisch Nederlands representeert.

In wat volgt, zal ik een digitaal compilatiecorpus historisch Nederlands (verder kortweg compilatiecorpus) voorstellen dat een antwoord wil bieden op de bovenstaande praktische en methodologische problemen. Het corpus is oorspronkelijk samengesteld in het kader van mijn doctoraatsonderzoek naar woordvolgordepatronen in tweeledige werkwoordclusters (Coussé 2008). Ik zal de lezer stap voor stap de selectiecriteria voorleggen die gehanteerd zijn bij de

samenstelling van het corpus. Deze bijdrage kan dan ook in de eerste plaats gelezen worden als een theoretische reflectie over de samenstelling van een — zo ideaal mogelijk — longitudinaal corpus historisch Nederlands. Daarnaast kan deze bijdrage ook gebruikt worden als een praktische handleiding bij het compilatiecorpus dat gratis te downloaden is via de tst-centrale (www.tst.inl.nl). Door het nieuw samengestelde corpus online aan te bieden wil ik andere historische taalkundigen de kans bieden onmiddellijk aan de slag te gaan met een betrouwbaar corpus historische teksten zonder zelf eerst al te veel tijd te verliezen met het vele praktische en methodologische voorbereidende werk dat komt kijken bij het samenstellen van een eigen corpus. Ik wil evenwel benadrukken dat het corpus dat in deze bijdrage wordt voorgesteld slechts een methodologisch verantwoorde compilatie is van reeds uitgeven historische bronnen in digitale en papieren vorm. Het samenstellen van een echt nieuw corpus dat de beschikbare digitale bronnen systematisch aanvult met het vele bronmateriaal dat nog in de archieven ligt, blijft werk voor een veel groter opgezet corpusproject in de toekomst.

2. Methodologische uitdagingen

Het samenstellen van een corpus historische teksten dat de hele geschiedenis van het Nederlands bestrijkt van de vroegste doorlopende teksten uit de dertiende eeuw tot het moderne Standaardnederlands is beslist geen sinecure. In wat volgt, zal ik ingaan op een tweetal ontwikkelingen in de geschiedenis van het Nederlands die het bijzonder lastig maken een longitudinaal corpus te compileren dat de hele vooropgestelde tijdspanne bestrijkt én dat voldoende evenwichtig samengesteld is qua taalgebruik en tekstgenre (vgl. ook Van der Horst 1997).

Om te beginnen zorgt de ontwikkeling van de geschreven standaardtaal ervoor dat het taalgebruik uit de middeleeuwse bronnen moeilijk te vergelijken is met de schrijftaal uit latere tijden (o.a. Van der Wal 1992, Van der Sijs 2004). In de middeleeuwen wordt de geschreven taal nog sterk gekleurd door het persoonlijke dialect van de auteur en/of de kopiist. In de loop van de zestiende en zeventiende eeuw wordt de schrijftaal onder impuls van de renaissance en de ontwikkelingen in de boekdrukkunst beregeld en gestandaardiseerd naar het voorbeeld van de klassieke talen (Van der Wal 1995). Hierbij is de geschreven standaardtaal vooral geënt op de Hollandse volkstaal, hoewel talige invloeden van zuidelijke en ook oostelijke immigranten niet onderschat mogen worden (voor discussie, zie Hendriks 1998, Van der Sijs 2004). We hebben dus te maken met een ontwikkeling van een weinig geüniformeerde schrijftaal waarbij de regionale herkomst van de auteur en/of de kopiist niet weg te denken is naar een meer gestandaardiseerde schrijftaal waarin regionale variatie zo veel mogelijk geëlimineerd is. Bij de selectie van de teksten voor het compilatiecorpus moet met die externe geschiedenis van het geschreven Nederlands voldoende rekening gehouden worden.

Naast het ontstaan van een geschreven standaardtaal zorgen ook enkele literaire ontwikkelingen tijdens de renaissance voor een breuk in het aanbod teksten uit de middeleeuwen en latere eeuwen. Zo gaat het Nederlands vanaf de renaissance het Latijn steeds meer vervangen als de prestigieuze taal voor wetenschap en cultuur (Van der Wal 1995). Dat betekent dat naast de tekstgenres die in de middeleeuwen al in de volkstaal geschreven werden (bv. epische gedichten, liederen, kluchten, receptenboeken, ambtelijke teksten) ook veeleer elitaire tekstgenres (bv. wetenschappelijke, politieke en religieuze traktaten, klassieke tragedies) steeds meer opgesteld worden in het Nederlands. Daarnaast zien we vanaf de renaissance naast de traditionele berijmde literatuur ook heel wat nieuwe tekstgenres ontstaan in prozavorm zoals de schelmenroman, het reisverhaal en later ook de briefroman, de historische roman, de novelle, het kortverhaal, enz. Beide literaire ontwikkelingen tonen aan dat het niet vanzelfsprekend is om één type literaire teksten te selecteren dat al acht eeuwen lang bestaat én dat al die tijd ook in het Nederlands werd geschreven.

3. Samenstelling van het compilatiecorpus

De geschetste ontwikkelingen in de geschiedenis van het Nederlands tonen dat het bijzonder moeilijk is — misschien zelfs in principe onmogelijk — om een longitudinaal corpus historisch Nederlands samen te stellen dat homogeen is op het vlak van taalgebruik en teksttype. Toch hoeft die methodologische uitdaging ons niet voor een volstrekte impasse te plaatsen. In wat volgt, zal ik uit de doeken doen hoe ik met de samenstelling van het Compilatiecorpus Historisch Nederlands een antwoord heb proberen te bieden op de geschetste complexe uitgangssituatie. Ik wil evenwel benadrukken dat de samenstelling van het compilatiecorpus slechts één mogelijke oplossing biedt voor de geschetste methodologische uitdagingen. Concreet heb ik in plaats van tevergeefs naar één homogeen corpus te streven voor de hele vooropgestelde tijdspanne twee deelcorpora verzameld die maar een stuk van de tijdspanne bestrijken en die elk één bepaald type taalgebruik weerspiegelen. Het eerste deelcorpus beoogt meer bepaald het regionaal gekleurde taalgebruik te representeren dat typisch is voor de geschreven taal uit de middeleeuwen. Het tweede deelcorpus richt zich dan weer op documenten in de standaardtaal vanaf het einde van de zestiende eeuw tot vandaag. In wat volgt, zal ik de precieze samenstelling van beide deelcorpora verder toelichten.

Het eerste deelcorpus wil zoals gezegd een representatieve steekproef vormen van de regionaal gekleurde schrijftaal uit de middeleeuwen. Ik herinner eraan dat er in de middeleeuwen nog geen sprake is van een verregaande standaardisatie in spelling, woordkeuze en grammatica zoals in de moderne schrijftaal, maar dat de geschreven taal nog sterk beïnvloed wordt door het persoonlijke dialect van de auteur. Voor een regionaal gediversifieerde steekproef is het nu van essentieel belang dat de verzamelde teksten betrouwbaar in de ruimte gelokaliseerd kunnen worden. Hier wringt helaas het schoentje bij

een aanzienlijk deel van de overgeleverde teksten uit de middeleeuwen. Veelal kennen we de identiteit van de middeleeuwse auteurs niet waardoor het gissen blijft naar de precieze herkomst van de schrijver of de plaats waar de tekst neergeschreven is. Toch zijn er wel degelijk een aantal middeleeuwse tekstgenres die betrouwbaar te lokaliseren zijn op basis van hun expliciete vermelding van de plaats en tijdstip van het schrijven in het document zelf (Van Loon 2002). Het gaat om egodocumenten zoals brieven of dagboeken en om ambtelijke documenten zoals oorkonden en processtukken die normaal gezien steeds een precieze vermelding van de tijd en plaats van het schrijven bevatten.

Het ligt voor de hand om van die goed gelokaliseerde bronnen de meest spontaan geformuleerde tekstgenres te kiezen zoals brieven of processtukken, waarvan we kunnen verwachten dat ze het dichtst aanleunen bij de spreektaal. Helaas zijn precies die spontane tekstgenres jarenlang relatief verwaarloosd bij de ontsluiting van de archieven door historici en taalkundigen. We mogen dan wel beschikken over enkele waardevolle uitgaven van de dagboeken of de briefwisseling van een aantal historische figuren, het aanbod echter is verre van groot én gediversifieerd genoeg om een representatieve steekproef van het taalgebruik uit de middeleeuwen samen te stellen.

Gelukkig is het met de ontsluiting van goed gelokaliseerde kanselarijdocumenten zoals oorkonden en statuten beter gesteld. Voor de hele middeleeuwse periode (en zelfs voor de daaropvolgende eeuwen tot het einde van het Ancien Régime in 1795) is de tekstproductie van de belangrijkste stedelijke kanselarijen uitgegeven in de vorm van grote tekstcollecties in zowel papieren als digitaal formaat. Het spreekt voor zich dat dergelijke rijke verzamelingen ambtelijke teksten een erg toegankelijke en gemakkelijk aan te boren bron vormen bij het samenstellen van het compilatiecorpus. Helaas zijn ambtelijke teksten dan weer in een formele en zelfs archaische stijl geschreven die mogelijk ver van het gewone taalgebruik uit die tijd staat. Toch kan de taalkundige studie van ambtelijke teksten ons wel degelijk iets leren over de geschiedenis van de Nederlandse taal. In onderzoek van Coussé (2010) is bijvoorbeeld aangetoond dat zelfs de erg frequente en bijzonder stereotiepe wending *zoals voorzeid is* onderhevig is aan diachrone veranderingsprocessen zoals grammaticalisatie. Soortgelijke bevindingen zijn door Marynissen (1999) en Boonen (2005) gepresenteerd voor de begin- en eindformules van oorkonden, waar heel wat meer vormvariatie optreedt dan doorgaans verwacht wordt van de stereotiepe delen van ambtelijke teksten.

Na heel wat wikken en wegen blijken ambtelijke teksten het best geschikt te zijn voor het invullen voor het eerste deelcorpus. Concreet heb ik voor originele ambtelijke teksten gekozen die zo goed mogelijk het lokale taalgebruik tussen de burgers uit de stad weerspiegelen. Het gaat om schepenbrieven, statuten en reglementen, lokale rechtspraak, verkoopsovereenkomsten, huurcontracten, schenkingen, huwelijkscontracten en wilsbeschikkingen. Ambtelijke communicatie tussen de lokale overheid en het hogere gezag zoals de graaf, hertog, koning of keizer komt niet in aanmerking omdat de herkomst van de

klerk veel onzekerder is en ook het taalgebruik meer bovengewestelijke kenmerken kan vertonen. Gezien de mogelijke dialectvariatie heb ik ambtelijke teksten gekozen uit drie centrale dialectstreken: enerzijds Vlaanderen en Brabant vanwege de vroege en rijke schriftelijke productie in de volkstaal en anderzijds Holland voor een optimale vergelijking met de latere standaardtaal (cf. *infra*).

Complementair met het corpus ambtelijke teksten, heb ik ook een deelcorpus met teksten verzameld die het meer gestandaardiseerde taalgebruik vanaf het einde van de zestiende eeuw moet weerspiegelen. De invulling van het tweede deelcorpus zorgt voor heel wat minder methodologische moeilijkheden dan het geval was bij het hoger beschreven ambtelijke deelcorpus. Om te beginnen hoeft bij het verzamelen van het tweede deelcorpus geen rekening meer gehouden te worden met mogelijke dialectvariatie in de teksten, gezien de regionale nivellering van het geschreven taalgebruik bij de ontwikkeling van de standaardtaal. Voor een optimale vergelijkbaarheid met de ambtelijke teksten heb ik evenwel uitsluitend teksten uit Holland gekozen, waar sinds het begin van de zeventiende eeuw bovendien een grote tekstproductie in druk ontstaan is. Daarnaast is het aanbod teksten die betrouwbaar in tijd en ruimte te plaatsen zijn niet beperkt tot egodocumenten en ambtelijke documenten, zoals dat in de middeleeuwen het geval was. Met de renaissance treden auteurs immers uit de anonimiteit zodat we een zicht krijgen op de herkomst van de schrijver en ook het moment van schrijven. Concreet zijn voor het tweede deelcorpus teksten geselecteerd waarvan bekend is dat de auteur een geboren en getogen Hollander is.

Bij de uiteindelijke selectie van teksten is ervoor geopteerd om in het tweede deelcorpus enkel prozateksten op te nemen. Ik herinner eraan dat er vanaf de renaissance naast de traditionele berijmde literatuur ook heel wat nieuwe tekstgenres ontstaan in prozavorm. Door enkel prozateksten toe te laten in het tweede deelcorpus sluit het taalgebruik van dat deelcorpus dichter aan bij het corpus ambtelijke teksten dat enkel prozateksten bevat. Op die manier verhoogt de homogeniteit van het gehele compilatiecorpus aanzienlijk en is het taalgebruik van beide deelcorpora tot op zekere hoogte te vergelijken. Om te benadrukken dat de prozateksten in het tweede deelcorpus een meer narratief — veelal zelfs een literair — karakter hebben dan de ambtelijke teksten in prozavorm zal ik in wat volgt naar het tweede deelcorpus verwijzen als het corpus narratieve teksten.

4. Compilatie van een evenwichtige steekproef

Nu de samenstelling van het compilatiecorpus uitgebreid verantwoord is, zal ik verder ingaan op de concrete invulling van het deelcorpus ambtelijke teksten en het deelcorpus narratieve teksten aan de hand van bestaande uitgegeven collecties historische teksten. Na een korte introductie van de geraadpleegde tekstverzamelingen zal ik uitvoerig ingaan op de selectie van de teksten uit die

grote corpora om zo tot een evenwichtige steekproef historische teksten te komen in overeenstemming met de methodologische keuzes uit de vorige twee paragrafen.

4.1 Deelcorpus ambtelijke teksten (1250–1799)

Bij de compilatie van het deelcorpus ambtelijke teksten kon ik voor de dertiende en veertiende eeuw putten uit twee gedigitaliseerde corpora met ambtelijke bronnen, nl. het eerder genoemde Corpus Gysseling en het Corpus Van Reenen-Mulder. Daarnaast bevat het deelcorpus ambtelijke teksten ook een selectie gescande teksten uit papieren tekstedities van de kanselarijbronnen van de grotere steden. In wat volgt, zal ik kort ingaan op de bijzonderheden van die drie broncollecties voor het deelcorpus ambtelijke teksten.

4.1.1 Corpus Gysseling (1250–1299)

Voor de oudste teksten in het ambtelijke deelcorpus heb ik een beroep kunnen doen op het Corpus van Middelnederlandse Teksten, een exhaustieve verzameling van alle overgeleverde literaire en ambtelijke teksten tot en met het jaar 1300. Die enorme tekstcollectie is in de jaren zeventig samengesteld door Maurits Gysseling, en staat daarom ook bekend als het Corpus Gysseling. Meer details over de precieze samenstelling van het corpus en over de manier waarop de teksten getranscribeerd werden, zijn na te lezen in Gysseling (1977). De oorspronkelijke papieren uitgave van het dertiende-eeuwse corpus is later door het Instituut voor het Nederlandse Lexicologie te Leiden gedigitaliseerd. Meer informatie over dat digitaliseringsproject, waarin het corpus ook van lemmatisering en woordsoortinformatie voorzien werd, is terug te vinden bij Pijnenburg en Schoonheim (1998). De digitale editie van het Corpus Gysseling is voor het grote publiek toegankelijk via de gebruiksvriendelijke interface van de al genoemde cd-rom Middelnederlands. Voor de samenstelling van het compilatiecorpus kon ik echter rechtstreeks gebruik maken van de achterliggende tekstbestanden die door het INL verstrekt zijn in het najaar van 2004.

4.1.2 Corpus Van Reenen-Mulder (1300–1399)

De ambtelijke teksten voor de veertiende eeuw zijn afkomstig uit het Corpus Veertiende-eeuwse Middelnederlandse Oorkonden, dat verzameld en gedigitaliseerd werd aan de Vrije Universiteit Amsterdam. Die oorkondeverzameling staat beter bekend als het Corpus Van Reenen-Mulder, genoemd naar de vroegste samenstellers van het corpus (Van Reenen en Mulder 1993). Later is het corpus verder aangevuld met meer zuidelijk tekstmateriaal in samenwerking met de Universiteit Gent. In tegenstelling tot het Corpus Gysseling bevat het Corpus Van Reenen-Mulder slechts een selectie van alle overgeleverde ambtelijke teksten uit de veertiende eeuw, aangezien de productie van ambte-

lijke teksten in de volkstaal tegen die tijd al te omvangrijk geworden is voor een integrale uitgave. We hebben dus te maken met een gecontroleerde steekproef, waarbij de veertiende-eeuwse ambtelijke teksten zoveel mogelijk in tijd en ruimte gespreid zijn. De digitale transcriptie van de ambtelijke teksten is net zoals de digitale editie van het Corpus Gysseling verrijkt met lemmatisering en woordsoortinformatie. In afwachting van de publieke release van het corpus, kon ik voor de samenstelling van het compilatiecorpus gebruik maken van een voorlopige versie van het corpus die in het najaar van 2004 door Piet van Reenen ter beschikking is gesteld.

4.1.3 Geschiedkundige rechtsbronnen (1400–1799)

Voor de periode vanaf de vijftiende eeuw zijn helaas geen digitale corpora met ambtelijke teksten beschikbaar, zodat ik voor het compilatiecorpus een beroep heb moeten doen op papieren tekstedities van de kanselarijbronnen van de grotere steden. Die tekstedities zijn tot stand gekomen omstreeks het einde van de negentiende eeuw in het kader van de grootschalige ontsluiting van de ambtelijke bronnen uit de stedelijke archieven door historici uit Nederland en België. Kenmerkend voor de geschiedkundige tekstedities is dat er slechts zelden commentaar gegeven wordt over hoe het tekstmateriaal verzameld is, over de manier waarop de handschriften precies getranscribeerd zijn en over de werkwijze waarop eventuele onduidelijke passages zijn opgelost. Soms kan men wel een voetnoot in de tekstuutgave aantreffen waarin melding gemaakt worden van een onleesbaar woord door bijvoorbeeld een watervlek, maar in tegenstelling tot het Corpus Gysseling en het Corpus Van Reenen-Mulder kan men bezwaarlijk van een echt diplomatisch verantwoorde uitgave spreken.

Een aantal van de tekstedities documenteren niet alleen de middeleeuwse rechtsgeschiedenis maar bieden ook een overzicht van de stadsadministratie vanaf de middeleeuwen tot aan het einde van het Ancien Régime (in de Nederlanden tot 1795). Ik heb van de gelegenheid gebruik gemaakt om het corpus ambtelijke teksten, dat in principe bedoeld is om de regionale variatie in het middeleeuwse taalgebruik te representeren, waar mogelijk aan te vullen met jongere ambtelijke documenten zoals notariële akten, stadsreglementen en officiële brieven uit de zeventiende en achttiende eeuw. Dergelijk continu aanbod ambtelijke teksten maakt het mogelijk om taalkundige tendensen ononderbroken te volgen binnen een homogeen tekstgenre vanaf de dertiende tot de achttiende eeuw in de dialectregio's Vlaanderen, Brabant en Holland.

4.1.4 Compilatie van de bronnen tot een evenwichtige steekproef

Nu de bijzonderheden van de drie gebruikte bronnen voor het deelcorpus ambtelijke teksten kort zijn voorgesteld, kan in meer detail ingegaan worden op de precieze compilatie van het deelcorpus tot een verantwoorde steekproef van het taalgebruik uit de middeleeuwen en daarna. In wat voorafging, heb ik

geargumenteerd dat het deelcorpus ambtelijke teksten de regionale variatie moet weerspiegelen uit de dialectstreken Vlaanderen, Brabant en Holland. Ik heb dan ook in de gebruikte bronnenverzamelingen enkel teksten geselecteerd uit die drie dialectstreken. Aangezien er in de bronnenverzamelingen een vrij grote spreiding is van de teksten over de verschillende steden heb ik ernaar gestreefd teksten te verzamelen uit een vijftal steden per regio om zo tot een nog betere regionale dekking van het deelcorpus te komen. Concreet zijn de Vlaamse teksten uit de steden Brugge, Ieper, Kortrijk, Gent en Oudenaarde afkomstig; de Brabantse teksten komen uit Brussel, Leuven, Mechelen, Antwerpen en Breda; en de Hollandse teksten ten slotte komen uit Dordrecht, Amsterdam, Haarlem, Gouda en Leiden. De keuze voor bovenstaande steden is grotendeels bepaald door het aanbod teksten in de geraadpleegde bronnen. Ik heb ernaar gestreefd voor elk van de vijftien gekozen steden ongeveer tweeduizend vijfhonderd woorden tekstmateriaal per tijdsdoorsnede van vijftwintig jaar te selecteren. In tabel 1 wordt een overzicht gegeven van het aantal woorden dat verzameld is per tijdsdoorsnede en per stad apart.

Hieruit blijkt dat het aardig gelukt is om een mooie spreiding van de teksten te bereiken op regionaal vlak en zelfs ook op stadsniveau in het deelcorpus ambtelijke teksten. Helaas vertoont de tabel ook nogal wat witte plekken door cellen die leeg gebleven zijn. In wat volgt, zal ik een aantal van die leemtes in de tabel proberen te motiveren.

Om te beginnen blijken de verzamelde teksten uit de dertiende eeuw slechts uit vier steden afkomstig te zijn, nl. Brugge, Gent, Mechelen en Dordrecht. Die tendens kan in verband gebracht worden met de ongelijkmatige verschriftelijking van het Nederlandse taalgebied (Burgers 1995). De productie van ambtelijke teksten in de volkstaal is meer bepaald het vroegst van start gegaan in de Vlaamse steden. Hierdoor was het geen probleem om in het Corpus Gysseling de beoogde hoeveelheid tekstmateriaal te verzamelen voor grote steden als Brugge en Gent. De schrijftraditie in de volkstaal begint later in Brabant en Holland, waardoor er in het Corpus Gysseling enkel vanaf het laatste kwart van de dertiende eeuw voldoende teksten beschikbaar zijn voor de steden Mechelen en Dordrecht. Om de dataschaarste in de dertiende eeuw enigszins op te vangen, heb ik waar mogelijk ook meer tekstmateriaal verzameld in de vier steden dan de vooropgestelde tweeduizend vijfhonderd woorden. Rekening houdend met het ongelijke regionale aanbod in het Corpus Gysseling valt de regionale spreiding van de ambtelijke teksten voor de dertiende eeuw vrij evenwichtig uit in het deelcorpus.

Daarnaast toont de tabel een opvallend tekort aan Brabantse teksten, in het bijzonder vanaf de vijftiende eeuw. Die tekstschaarste kan toegeschreven worden aan een gebrek aan beschikbare papieren tekstedities voor de Brabantse steden. De enkele beschikbare tekstcollecties blijken bovendien hoofdzakelijk de oudste oorkonden te herbergen, waardoor de geschiedkundige tekstverzamelingen voor Brabant weinig nieuwe teksten aandragen naast wat al beschikbaar was via het Corpus Gysseling en het Corpus Van Reenen-Mulder.

Tabel 1: Invulling van het deelcorpus ambtelijke teksten (n = 393 332 woorden)

| | Vlaanderen | | | Brabant | | | Holland | | | Totaal | |
|---------------|--------------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|
| | Brug. Ieper | Gent | Kortr. Oud. | Antw. Breda | Bruss. Leuv. | Mech. | Amst. Dordr. | Goud. Haarl. | Leiden | | |
| 1250-74 | 22829 | 89 | 3770 | | | 2863 | | | | 29551 | |
| 1275-99 | 20255 | 16319 | 699 | | 601 | 15242 | 11479 | | 130 | 64725 | |
| 1300-24 | 2153 | 234 | 531 | 2428 | 650 | 1932 | 3894 | 239 | 425 | 686 | |
| 1325-49 | 1790 | 1734 | 1002 | | 1172 | 1948 | 360 | 1377 | 3230 | 662 | |
| 1350-74 | 1910 | 1377 | 1717 | 400 | 1769 | 1464 | 1879 | 1921 | 2858 | 1340 | |
| 1375-99 | 1801 | 2442 | 1054 | 1415 | 2507 | 1519 | 2786 | 2662 | 2909 | 2891 | |
| 1400-24 | 1389 | 1528 | 2836 | 82 | 1328 | 1420 | 1543 | 2735 | 2975 | 2797 | |
| 1425-49 | 2944 | 1008 | 1977 | | 2860 | 1664 | 2767 | 2346 | 444 | 3312 | |
| 1450-74 | 3159 | 309 | 2368 | | 3083 | | 2301 | 2730 | 2063 | 1991 | |
| 1475-99 | 2740 | 423 | 3002 | | 2104 | | 2547 | 1558 | 1682 | 2304 | |
| 1500-24 | 2045 | 3205 | 2864 | 761 | | | 2845 | 2334 | 2268 | 2805 | |
| 1525-49 | 1601 | 1618 | 1280 | 997 | | | | 2792 | 1919 | 2428 | |
| 1550-74 | 1624 | 592 | 1385 | 1349 | | 309 | | 2885 | 1981 | 2704 | |
| 1575-99 | 2422 | 2497 | 178 | 890 | | | | 2197 | | 2168 | |
| 1600-24 | 228 | 1270 | 1456 | | | | | | | 1728 | |
| 1625-49 | 1233 | 122 | 2033 | | | 998 | 2890 | 2886 | | 2324 | |
| 1650-74 | | 2763 | 2334 | | | 3607 | 2557 | | | 2254 | |
| 1675-99 | 790 | | | | | | | | | 2228 | |
| 1700-24 | 825 | | 442 | | | | | | | 2414 | |
| 1725-49 | 1265 | 2400 | 351 | | | 4812 | | | | 1498 | |
| 1750-74 | 2877 | 1197 | 215 | | | | | | | 2880 | |
| 1775-99 | | | 380 | | | | | | | 2313 | |
| Totaal | 75880 | 22311 | 51156 | 8322 | 15473 | 15462 | 25361 | 40910 | 22568 | 23659 | 393332 |

Een vergelijkbaar probleem van dataschaarste treedt ook vanaf de zeventiende eeuw voor Holland op. Het merendeel van de papieren tekstedities voor de Hollandse steden loopt maar tot 1600, zodat slechts voor Amsterdam en Leiden ook jongere teksten voorhanden zijn.

Al bij al biedt het deelcorpus ambtelijke teksten een steekproef van ambtelijke teksten verspreid over vijftien steden uit drie verschillende dialectregio's vanaf de tweede helft van de dertiende eeuw tot het einde van de achttiende eeuw. Hoewel zo veel mogelijk getracht is om evenveel tekstmateriaal uit de vijftien steden te verzamelen voor elke tijdsdoorsnede, lieten het huidige aanbod uitgegeven ambtelijke teksten niet toe om een steekproef samen te stellen die dialectologisch onderzoek toe moeten laten tot op stadsniveau. Hiervoor zal bijkomend excerpeerwerk nodig zijn van ambtelijke documenten die dateren van na de vijftiende eeuw in de stadsarchieven.

4.2 Deelcorpus narratieve teksten (1575–2000)

Terwijl voor het deelcorpus ambtelijke teksten gebruik gemaakt moest worden van verschillende bronnencollecties is het deelcorpus met narratieve teksten voornamelijk afkomstig van de Digitale Bibliotheek voor de Nederlandse Letteren (zie www.dbnl.org). Die online bibliotheek bevat een schat aan gedigitaliseerde literaire teksten vanaf de vroegste bronnen in het Nederlands tot vandaag, aangevuld met secundaire literatuur en biografische informatie. Uit dat ruime aanbod heb ik enkel narratieve teksten verzameld waarvan bekend is dat de auteur een geboren en getogen Hollander is. Dat houdt in dat de oudste teksten van het deelcorpus narratieve teksten pas vanaf het einde van de zestiende eeuw dateren, aangezien de oudere narratieve teksten in de dbnl doorgaans anoniem overgeleverd zijn.

De precieze keuze van het type narratieve teksten hangt in sterke mate af van het aanbod in de Digitale Bibliotheek voor de Nederlandse Letteren. In de zeventiende eeuw zijn de narratieve teksten in de dbnl vooral beperkt tot religieuze, filosofische, geschiedkundige en politieke traktaten. Meer literaire werken zoals kluchten of tragedies zijn nog in hoofdzaak berijmd en komen dus niet in aanmerking voor het deelcorpus narratieve teksten. Om de relatieve schaarste van narratieve werken in de zeventiende eeuw op te vangen heb ik ervoor geopteerd ook de inleidingen bij berijmden werken te exciperen. Die inleidingen van de hand van de auteur of de uitgever zijn doorgaans slechts korte narratieve teksten, maar ze dragen bij tot een verdere diversificatie van de narratieve teksten in de zeventiende eeuw. Met de verdere ontwikkeling van de Nederlandse literatuur worden meer tekstgenres geschikt voor het deelcorpus narratieve teksten. In de latere tijdsdoorsneden zijn naast traktaten en inleidingen ook meer literaire genres als de schelmenroman, het reisverslag, de historische roman, de novelle en het kortverhaal opgenomen in het deelcorpus.

Voor elke tijdsdoorsnede van vijftwintig jaar heb ik ernaar gestreefd om een vijftal narratieve teksten van verschillende auteurs te verzamelen van elk tweeduizend vijfhonderd woorden lang. In tegenstelling tot de meeste ambtelijke teksten overschrijden de narratieve teksten doorgaans dat vooropgestelde streefcijfer ruimschoots. Met het oog op een evenwichtige vertegenwoordiging van de verschillende auteurs in het corpus, heb ik ervoor gekozen om toch vast te houden aan de invulling van elke cel door tweeduizend vijfhonderd woorden in plaats van al het beschikbare materiaal op te nemen. Enkel wanneer er binnen één tijdsdoorsnede een schaarste aan bronnen dreigde, is van dit principe afgeweken door meer materiaal voor enkele auteurs te verzamelen. In de appendix is een overzicht gegeven van de gekozen auteurs, de titel van de narratieve teksten, het tekstgenre van die teksten en ten slotte de grootte van de fragmenten die in het deelcorpus geëxcerpeerd zijn.

5. Uniforme vormgeving van het compilatiecorpus

Nu de precieze invulling van het compilatiecorpus grondig uit de doeken is gedaan, moet nog de nodige aandacht besteed worden aan de manier waarop de verzamelde teksten tot een uniform corpus omgevormd zijn. Ik zal achtereenvolgens bespreken hoe de uiteenlopende opmaak op het vlak van taalkundige annotaties en beschikbare metatagale informatie is geüniformeerd in de verzamelde steekproef.

Om te beginnen zijn de verzamelde teksten niet op een consistente manier van taalkundige annotaties voorzien in de verschillende bronnen. Zo bevatten de ambtelijke teksten uit het Corpus Gysseling en het Corpus Van Reenen-Mulder woordsoortinformatie en lemmatisering terwijl de gescande ambtelijke teksten en narratieve teksten van de dbnl helemaal niet taalkundig verrijkt zijn. Oorspronkelijk was het de bedoeling om de taalkundige annotatie van het Corpus Gysseling en het Corpus Van Reenen-Mulder te behouden in het nieuw samengestelde compilatiecorpus. Dat betekent dat er gestreefd zou moeten worden naar een uniforme opmaak van de taalkundige informatie in de tekstbestanden voor een optimale doorzoekbaarheid in het hele compilatiecorpus. Het maken van uniform vormgegeven annotaties bleek al heel gauw een erg delicate klus. Zo zijn de codes voor de taalkundige verrijking in beide bronbestanden op een afwijkende manier aan de woorden gehecht (in het Corpus Gysseling door zogenaamde vishaken rond het woord en in het Corpus Van Reenen-Mulder door een liggend streepje achter het woord of in de meest recente bestanden in de complexe XML-opmaak). Daarnaast wijken de gebruikte taalkundige codes in beide bronbestanden soms op minieme punten van elkaar af. Beide verschillen in opmaak van de taalkundige annotaties zijn niet onoverkomelijk, maar het lijkt wenselijker dat de samenstellers van beide corpora samen een standaard proberen te bereiken, eventueel in het kader van een overkoepelend corpusproject dat beide corpora tot één verzameling ambtelijke teksten samenbrengt en aanvult met jongere ambtelijke documenten. Bo-

vendien zou het behouden van de taalkundige annotaties voor de dertiende- en veertiende-eeuwse ambtelijke teksten voor de bijkomende complicatie gezorgd hebben dat het vroegste deel van het compilatiecorpus niet hetzelfde formaat had als de rest van het corpus (dat het leeuwendeel van het tekstmateriaal uitmaakt). Ik heb dus uiteindelijk het vroegste deel van het compilatiecorpus van zijn annotaties ontdaan in overeenstemming met de latere niet-verrijkte delen van het corpus. Het verwijderen van het kluwen van codes en haakjes uit het corpus heeft ook als voordeel dat de leesbaarheid van de teksten aanzienlijk stijgt.

Niet alleen op het vlak van taalkundige annotaties, maar ook wat de beschikbare metatalige informatie betreft, wijken de teksten uit de verschillende bronnen aanzienlijk van elkaar af. Zo bevatten de digitale tekstbestanden afkomstig uit het Corpus Gysseling en het Corpus Van Reenen-Mulder bijzonder veel details over het getranscribeerde manuscript zelf (het gebruik van speciale tekens of afkortingen in de tekst, het voorkomen van onleesbare passages, het begin van een nieuwe regel of bladzijde, afwijkingen in het handschrift die wijzen op verschillende kopiïsten). In de gescande oorkonden echter vinden we slechts hier en daar een sporadische voetnoot met wat beknopte informatie over een onleesbare passage als bijvoorbeeld het gevolg van een watervlek. In de Digitale Bibliotheek voor de Nederlandse Letteren is ten slotte veel zorg gehecht aan het adequaat weergeven van titels, paragrafen en paginanummers uit gedrukte publicaties.

Uit dat bijzonder heterogeen aanbod metatalige informatie heb ik de gegevens gedestilleerd die beschikbaar waren voor alle teksten in het compilatiecorpus. Om te beginnen heb ik informatie over de herkomst van een tekst en het jaar waarin die tekst geschreven is uit de tekstbestanden gehaald en op een uniforme manier verwerkt in de bestandsnaam van elke tekst. Die documentnaam bestaat achtereenvolgens uit de stad waar de tekst neergeschreven is, het jaartal van compilatie en een volgnummer die teksten van elkaar onderscheidt die op dezelfde tijd en plaats geschreven zijn (bv. `amsterdam_1349_1.txt`). Daarnaast zijn woorden of passages die door de uitgever van een teksteditie op één of andere manier als onleesbaar gemarkeerd zijn systematisch tussen rechte haakjes geplaatst (bv. in `[orconde] desen brieue`). De aanduiding van een nieuwe paragraaf is systematisch weergegeven door middel van een tabteken zodat de tekst netjes visueel gestructureerd wordt. Alle andere metatalige informatie is uit de tekstbestanden verwijderd zodat uiteindelijk een platte tekst overblijft. In die grote opschoonoperatie is ook alle interpunctie systematisch uit de tekstbestanden verwijderd. De oorspronkelijk motivatie hiervoor was dat leestekens in het Middelnederlands op een heel andere manier in de lopende tekst zijn aangebracht dan vandaag de dag het geval is. Bovendien vergemakkelijkt de afwezigheid van interpunctie het zoeken naar aaneengesloten reeksen van woorden zonder hierbij rekening te hoeven houden met doorbrekende leestekens. De keuze is achteraf gezien nogal ongelukkig, aangezien leestekens in een digitale tekst vrij eenvoudig zelf door de gebruiker ver-

wijderd kunnen worden door middel een *tokenizer* programma. Bovendien heeft het verwijderen van interpunctie ook de algemene leesbaarheid van de tekstbestanden tot op zekere hoogte aangetast, hoewel het gebruik van hoofdletters dat nadeel toch grotendeels weer goedmaakt.

Ten slotte zijn er in de platte tekst van ambtelijke teksten annotaties toegevoegd die het clichématige begin en einde van een oorkonde onderscheiden van het meer verhalende middendeel. Door de verschillende onderdelen van een oorkonde te markeren, kan een onderzoeker ervoor kiezen om de meest stereotiepe stukken uit van de ambtelijke tekstproductie uit het compilatiecorpus te weren (cf. Coussé 2008).

In de volgende figuur is de uiteindelijke opmaak van de tekstbestanden geïllustreerd met een veertiende-eeuwse oorkonde uit Amsterdam (amsterdam_1349_1.txt).

```
<begin>
Wi jacob gherits soen ende jacob ghisetgijns soen scepene in
Aemstelredamme orconden ende kennen
</begin>
<main>
dat peter hilmers soen gheliede voer ons dat [hi] scoudich is heyndric
iacobs soen tien scellinghe hollants an comans ghelt iarlic ser rente
staende tot eweliken daghen op siin huus ende op erue daer hi nv
inwoent gheleghen an die steghe diemen ter [olen] gaet tusschen claes
hauics erue op die ene zide ende peter leyts erue op die ander zide
tebetalen dese rente alle iare op den meye dach
</main>
<end>
in [orconde] desen brieue bezeghelt mit onsen zeghelen Ghegheuen inden
jare ons heren mccc neghen ende viertich des woensdaghes na sente
iacobs dach
</end>
```

6. Troeven en beperkingen van het compilatiecorpus

Om de bespreking van de samenstelling en vormgeving van het compilatiecorpus af te sluiten, wil ik kort de troeven en — helaas ook onvermijdelijk — de beperkingen van het corpus evalueren voor historisch onderzoek van het Nederlands.

De grootste troef van het compilatiecorpus ligt mijns inziens in de zorgvuldige selectie van de teksten volgens een aantal welgedefinieerde criteria zoals de regionale herkomst en de datering van de teksten. Zo is er in het deelcorpus ambtelijke teksten systematisch naar gestreefd een evenwichtige steekproef samen te stellen van originele, lokale ambtelijke teksten uit de drie dialectstreken Vlaanderen, Brabant en Holland. Binnen elke tijdsdoorsnede van vijftienvintig jaar is getracht om steeds ongeveer tweeduizend vijfhonderd woorden aan tekstmateriaal te verzamelen in vijftien verschillende steden voor

de tijdspanne van 1250 tot 1800. Het deelcorpus narratieve teksten beoogt dan weer het geschreven taalgebruik in Holland evenwichtig te representeren vanaf het einde van de zestiende eeuw tot vandaag. Per tijdsdoorsnede van vijftwintig jaar is hier gestreefd naar een gediversifieerde steekproef van tekstmateriaal van een vijftal verschillende schrijvers die in Holland geboren en getogen zijn. De evenwichtige samenstelling van het compilatiecorpus maakt het mogelijk om taalveranderingen ononderbroken in de tijd en ruimte te volgen vanaf het vroegste Nederlands tot vandaag. Om een zicht te geven op de precieze mogelijkheden van het compilatiecorpus zal ik in wat volgt kort enkele al bestaande casestudies aanhalen die met behulp van het compilatiecorpus zijn uitgevoerd.

Om te beginnen blijkt het compilatiecorpus bijzonder geschikt voor syntactisch onderzoek van constructies die een vrij hoge frequentie hebben in geschreven taalgebruik. Het compilatiecorpus is oorspronkelijk samengesteld in het kader van mijn doctoraatsonderzoek naar woordvolgordepatronen in tweeledige werkwoordclusters (Coussé 2008). Tweeledige werkwoordclusters hebben een vrij hoge incidentie in het compilatiecorpus, aangezien ze in het Nederlands in toenemende mate gebruikt worden voor het uitdrukken van de voltooide tijd (bv. *heeft geschreven, is gekomen*), het passief (bv. *wordt verkocht*) of de toekomstige tijd (bv. *zal geschieden*). Concreet konden 4327 attestaties van tweeledige clusters geëxcerpeerd worden in het deelcorpus ambtelijke teksten en 1681 attestaties in het deelcorpus narratieve teksten. Op basis van die overvloedige attestaties was het mogelijk om de evolutie van woordvolgordepatronen in tweeledige werkwoordclusters ononderbroken te traceren vanaf het vroege Middelnederlands tot het hedendaagse Nederlands tot op vijftwintig jaar nauwkeurig. Het compilatiecorpus bevatte bovendien ook voldoende attestaties om naast de werkwoordsvolgorde ook andere volgordepatronen in de zin te onderzoeken in tijdsdoorsneden van vijftig jaar (zie ook Coussé 2009) en daarnaast ook een zicht te geven op de grammaticalisatie van de tweeledige werkwoordclusters (zie ook Coussé 2006). Het valt te verwachten dat het compilatiecorpus even bruikbaar zal zijn voor syntactisch onderzoek van constructies met een vergelijkbare of zelfs hogere frequentie.

Daarnaast heeft het compilatiecorpus in zijn korte bestaan ook al zijn nut bewezen bij morfologisch onderzoek van historisch Nederlands. Zo legden De Vogelaer en Coussé (2008) op basis van het deelcorpus ambtelijke teksten de diachrone ontwikkeling van complexe persoonlijke meervoudspronomina (bv. *jullie, haarlieden*) naast de oorspronkelijke simplexvormen (bv. *wij, haar*) bloot voor het Middelnederlands. Met een relatief hoge frequentie van 6819 geattesteerde voornaamwoorden kon de verhouding tussen de complexe meervoudsvormen en de simplexvormen tot op vijftig jaar nauwkeurig onderzocht worden voor de drie dialectregio's Vlaanderen, Brabant en Holland afzonderlijk. Het ligt zeker binnen de mogelijkheden om de ontwikkeling van andere morfologische verschijnselen van een vergelijkbare frequentie te onderzoeken met behulp van het compilatiecorpus. Ik wil er evenwel aan herinneren dat het cor-

pus niet voorzien is van enige woordsoortinformatie of lemmatisering, iets wat morfologisch onderzoek aanzienlijk vergemakkelijkt. Daarnaast is ook enkel het oudste deel van het compilatiecorpus afkomstig van diplomatisch getranscribeerde bronnen, waardoor het corpus niet meteen geschikt is voor morfologisch onderzoek naar bijvoorbeeld casusverlies, waar een accurate transcriptie van de eindletters van woorden van het grootste belang is.

Zo ben ik geleidelijk tot de beperkingen van het compilatiecorpus gekomen. Aangezien het corpus samengesteld is met het oog op syntactisch onderzoek van de werkwoordgroep, is het daar uiteraard het beste op toegerust qua omvang en transcriptiedetail. Het spreekt voor zich dat het compilatiecorpus al gauw tegen zijn limieten aanloopt bij onderzoek naar minder frequente constructies, zoals de ditransitieve constructie. Bij dergelijk weinig voorkomende verschijnselen is het vooral van belang om een voldoende groot corpus teksten bij elkaar te verzamelen, eventueel ten koste van de strikte criteria die voor het compilatiecorpus zijn gehanteerd. Daarnaast is het compilatiecorpus niet bijzonder geschikt voor klassiek dialectologisch onderzoek dat de geografische distributie van taalverschijnselen tot op het niveau van de stad wil blootleggen. Bij de samenstelling van het ambtelijke deelcorpus bleek immers niet voor elke stad voldoende materiaal ter beschikking te zijn voor de hele tijdspanne tot het einde van de achttiende eeuw.

7. Aanbevelingen voor de toekomst

Een opsomming van de beperkingen van het compilatiecorpus hoeft niet noodzakelijk tot pessimisme te leiden, maar kan evengoed gelezen worden als een aanbeveling voor toekomstige corpusprojecten. Het spreekt voor zich dat de bestaande ambtelijke corpora van de dertiende en veertiende eeuw in de toekomst aangevuld zullen moeten worden met even geografisch gediversifieerd en kwalitatief getranscribeerd en geannoteerd materiaal uit latere periodes. Daarnaast zou het ook interessant zijn om de Digitale Bibliotheek voor de Nederlandse Letteren beter toegankelijk te maken voor taalkundig onderzoek. Op dit moment moet de historisch taalkundige het erg diverse tekstmateriaal op een nogal omslachtige manier doorzoeken op geschikte corpusteksten, zelf bibliografische gegevens van de auteurs zoeken en ten slotte de gekozen teksten één voor één downloaden. Ten slotte is het de ultieme wens van iedere historisch taalkundige om in de toekomst gebruik te kunnen maken van een gediversifieerd corpus met historische teksten van de vroegste bronnen tot vandaag dat uniform is vormgegeven en op verschillende taalkundige niveaus is geannoteerd. Alleen valt bij een dergelijk onderneming te verwachten dat de methodologische uitdagingen die in paragraaf twee geschetst zijn voor moeilijkheden zullen blijven zorgen: er is en blijft een discrepantie in het aanbod geschreven taal in de geschiedenis van het Nederlands en ook het concept van de geschreven taal zelf is veranderd door de tijd heen.

Bibliografie

- Boonen, U.** 2005. De begin- en slotformules in Utrechtse oorkonden uit de dertiende en veertiende eeuw: een vergelijking van Middelnederlandse en Latijnse formuleringen. *Neerlandistiek.nl* 05(06): 1-55.
- Burgers, J.** 1995. Over de lokalisering van Middelnederlandse ambtelijke bescheiden. *Taal en Tongval* themanummer 8 *Historische dialectologie*: 139-164.
- Coussé, E.** 2006. De historische wortels van volgordevariatie in het *hebben*-perfectum. *Taal en Tongval* 58: 250-277.
- Coussé, E.** 2007. Digitale bronnen voor taalhistorisch onderzoek van het Nederlands. *Nederlandse Taalkunde* 12: 275-279.
- Coussé, E.** 2008. *Motivaties voor volgordevariatie. Een diachrone studie van werkwoordsvolgorde in het Nederlands*. Proefschrift. Gent: Universiteit Gent.
- Coussé, E.** 2009. Focus, complexiteit en extrapositie. Over de veranderende woordvolgorde in het Nederlands. *Neerlandistiek.nl* 09(04): 1-31.
- Coussé, E.** 2010. Grammaticalisatie in de ambtelijke formule 'zoals voorzeid is'. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 126: 20-33.
- De Vogelaer, G. en E. Coussé.** 2008. De kracht van disambiguering: nieuwe meervoudspronomen van het Middelnederlands tot nu. *Taal en Tongval* themanummer 21 *Dialectgeografie en interne factoren*: 13-35.
- Gysseling, M.** 1977. *Corpus van Middelnederlandse teksten (tot en met het jaar 1300)*. 's-Gravenhage: Nijhoff.
- Hendriks, J.B.** 1998. *Immigration and Linguistic Change. A Socio-cultural Linguistic Study of the Effect of German and Southern Dutch Immigration on the Development of the Northern Dutch Vernacular in the 16th/17th-Century Holland*. Proefschrift. Madison: University of Wisconsin.
- Marynissen, A.** 1999. ... allen dengenen die dese letteren sien selen / selen sien ende horen lezen ... Over volgordevariatie in de werkwoordelijke eindgroep in de Middelnederlandse bijzin. *Taal en Tongval* themanummer 12 *De verschriftelijking van het Nederlands*: 136-158.
- Pijnenburg, W.J.J. en T.H. Schoonheim.** 1998. De geschiedenis van een project. Schoonheim, T.H. en Th.P.F. Wortel (Reds.). 1998. *Een samenleving verwoord. Het Vroegmiddelnederlands Woordenboek*: 9-27. Den Haag: Sdu Uitgevers.
- Van der Horst, J.M.** 1997. Over en naar aanleiding van Zuid-Nederlandse doorbrekingen. Van Santen, A. en M.J. Van der Wal (Reds.). 1997. *Taal in tijd en ruimte*: 299-307. Leiden: SNL.
- Van der Sijs, N.** 2004. *Taal als mensenwerk, de geschiedenis van het ontstaan van het ABN*. Den Haag: Sdu Uitgevers.
- Van der Wal, M.J.** 1992. *Geschiedenis van het Nederlands*. Utrecht: Het Spectrum.
- Van der Wal, M.J.** 1995. *De moedertaal centraal. Standaardisatie-aspecten in de Nederlanden omstreeks 1650*. Den Haag: Sdu Uitgevers.
- Van Loon, J.** 2002. Op zoek naar de ideale tekst. Vanhoutte, E. (Red.). 2002. *Talig erfgoed. De zuidelijke Nederlanden in de 14de eeuw*: 73-90. Gent: KANTL.
- Van Reenen, P.Th. en M. Mulder.** 1993. Een gegevensbank van 14de-eeuwse Middelnederlandse dialecten op de computer. *Lexikos* 3: 259-281.

Appendix: Invulling van het deelcorpus narratieve teksten (n = 214 338 woorden)

| Periode | Auteur | Jaar | Titel | Genre | Tot. |
|---------|---|------------------------------|---|--|----------------------------|
| 1572-99 | Jacobsz., Wouter Coornhert, Dirk Van Hout, Jan | 1572 1585 1596 | <i>Dagboek Broeder Wouter Jacobsz.</i> <i>Zedenkunst dat is wellevenskunste</i> <i>Loterijspel</i> | dagboek traktaat voorwoord | 4334 5652 2261 |
| 1600-24 | De Groot, Hugo | 1613 | <i>Der heeren Staten van Hollandt ende West-Vrieslandt godts- diensticheyt</i> | traktaat | 5229 |
| | Orlers, Jan J. Bredero, Gerbrand A. Van Hogendorp, Gijsbrecht | 1614 1617 1617 | <i>Beschrijvinge der stad Leyden</i> <i>Den Spaanschen Brabander</i> <i>Truer-spel van de moordt, begaen aen Wilhelm, prince van Oraengien</i> | traktaat voorwoord voorwoord | 5607 2300 1125 |
| | Vander Plasse, Cornelis L. Bredero, Gerbrand A. Van de Venne, Adriaen Vander Plasse, Cornelis L. | 1619 1622 1622 1622 | <i>De klucht van den molenaar (van Bredero)</i> <i>Groot lied-boeck</i> <i>Tafereel van sinne-mal</i> <i>Groot lied-boeck (van Bredero)</i> | voorwoord voorwoord voorwoord voorwoord | 1084 1378 248 285 |
| 1625-49 | Vos, Jan Hooft, Pieter C. Coster, Samuel Six, Jan | 1641 1642 1648 1648 | <i>Aran of Titus</i> <i>Nederlandsche Historien. Het Leids beleg en ontzet, 1574</i> <i>De ses eerste vertoningen op de eeuwige vrede</i> <i>Medea</i> | voorwoord kroniek voorwoord voorwoord | 1308 7444 774 679 |
| 1650-74 | Huygens, Constantijn Vos, Jan Meyer, Lodewijk | 1667 1667 1668 | <i>Zee-straet</i> <i>Medea</i> <i>Verloofde koninksbruidt</i> | traktaat voorwoord voorwoord | 3298 3604 3738 |
| 1675-99 | Brandt, Geeraardt Heinsius, Nicolaas Rotgans, Lukas | 1682 1695 1698 | <i>Het leven van Joost van den Vondel</i> <i>Den vernaketyken avonturier</i> <i>Wilhelm de Derde</i> | biografie roman voorwoord | 6046 4004 743 |
| 1700-24 | Van Hoogstraten, David Sewel, Willem Alewyn, Abraham | 1700 1708 1714 | <i>Aenmerkingen over de geslachten der zelfstandige naemwoorden</i> <i>Nederduytsche spraakkonst</i> <i>Beslikte Swaanitie en drooge Fobert</i> | voorwoord voorwoord voorwoord | 3124 1568 395 |

| | | | | | |
|---------|---|--------------------------------------|---|---|--------------------------------------|
| 1725-49 | Langendijk, Pieter Alewyn, Abraham Bidloo, Lambert Ten Kate, Lambert | 1714 1719 1720 1723 | Het wederzijds huwelyksbedrog De Puiterveense helleveeg Pamphoëticon Bataarum Aenleiding tot het verhevene deel der Nederduitsche sprake | voorwoord voorwoord voorwoord voorwoord | 599 512 787 6338 |
| 1725-49 | Hoogvliet, Arnold Poot, Hubert | 1728 1728 | Abraham, de aartsouder Gedichten | voorwoord voorwoord | 2015 2264 |
| 1750-74 | Huydecooper, Balthazar Kerstman, Franciscus L. Van Winter, Nicolaas Van Alphen, Daniel | 1730 1756 1769 1772 | Proeve van taal- en dichtkunde Zeldzame levens-gevallen van J.C. Wyerman De jaargetyden Levenbericht van Jan Wagenaar | voorwoord biografie voorwoord grafrede | 2174 9006 233 4167 |
| 1775-99 | Emmery, Willem Corver, Marten Kinker, Johannes Paape, Gerrit Fokke, Arend S. | 1782 1786 1788 1789 1792 | Onderwijs voor kinderen Tooneel-aanteekeningen De post van den Helicon Het land der willekeurigen De moderne Helicon, een droom | voorwoord traktaat spectator reisverhaal satire | 3855 1257 2129 2570 2140 |
| 1800-24 | Bilderdijk, Willem Klijn, Hendrik Da Costa, Isaac Loosjes, Adriaan | 1807 1814 1823 1823 | De ziekte der geleerden Jan Fraderik Helmers, in eene redevoering uitgesproken Bezwaren tegen den geest der eeuw Het leven van Mauritz Lijnslager | voorwoord grafrede pamflet roman | 2742 6129 3568 3000 |
| 1825-49 | Drost, Aarnout Geel, Jacob Bosboom-Toussaint, Anna Kneppelhout, Johannes Van Koetsveld, Cornelis E. | 1832 1835 1840 1841 1843 | Hemingard van de Eikenterpen Gesprek op den Drachtenfels Het huis Lauernesse Studenten-Typen Schetsen uit de pastorij te Maastland | roman traktaat roman schets schets | 4060 2486 1671 2741 2464 |
| 1850-74 | Multatuli De Génesstet, Petrus A. Wolbers, J. Van Schaick, Cornelis Beets, Nicolaas | 1860 1861 1861 1866 1867 | Max Havelaar Over kinderpoëzy Geschiedenis van Suriname De Manja. Familie-tafereel uit het Surinaamsche volksleven Over kinderboeken. Gesprek met Crito | roman traktaat traktaat schets traktaat | 2009 1633 1355 2946 1976 |

| | | | | | |
|---------------|---|--------------------------------------|--|--|--------------------------------------|
| 1875-99 | Busken Huet, Coenraad Emants, Marcellus Vosmaer, Carel Van Eeden, Frederik Netscher, Frans | 1879 1879 1880 1885 1886 | <i>Het land van Rubens</i> <i>Een drietal novellen</i> <i>Amazona</i> <i>De kleine Johannes</i> <i>Studie's naar het naakt model</i> | reisverhaal novelle roman novelle schets | 3024 2675 4052 3627 4071 |
| 1900-24 | Couperus, Louis Van Booven, Henri Heijermans, Herman Van der Leeuw, Aart Van Looy, Jacobus | 1901 1904 1908 1908 1917 | <i>De boeken der kleine zielen</i> <i>Tropenwee</i> <i>Een wereldstad. Berlijnsche impressies en schetsen</i> <i>Sint-Veit</i> <i>Jaapje</i> | roman reisverhaal schets kortverhaal roman | 2007 3037 2947 4048 1559 |
| 1925-49 | Roland Holst, Adriaan Timmerman, Aegidius W. Terborgh, F.C. Reve, Gerhard Roland Holst, Adriaan | 1935 1938 1940 1949 1949 | <i>Nederland. Oorlogstuig</i> <i>Tim's herinneringen</i> <i>De condottiere</i> <i>De avonden</i> <i>Borrelpraat</i> | kortverhaal memoires kortverhaal roman kortverhaal | 1650 3236 3127 5215 802 |
| 1950-74 | Wolkers, Jan Presser, Jacob Vinkenoog, Simon Morriën, Adriaan Arends, Jan | 1963 1965 1965 1968 1974 | <i>Wespen</i> <i>Ondergang. De vervolging en verdelging van het Neder- landse jodendom</i> <i>Liefde. Zeventig dagen op ooghoogte</i> <i>Cryptogram</i> <i>Ik had een strohoed en een wandelstok</i> | kortverhaal traktaat dagboek anekdote kortverhaal | 3776 2732 1322 2396 1345 |
| 1975-99 | Mulisch, Harry Hermans, Willem F. Biesheuvel, Jacob M.A. Brakman, Willem Nootboom, Cees | 1975 1976 1983 1983 1991 | <i>Twee vrouwen</i> <i>De raadselachtige Multatuli</i> <i>Reis door mijn kamer</i> <i>Een waak in het kroos</i> <i>Het volgende verhaal</i> | roman biografie novelle essay roman | 1800 2345 1858 4066 2567 |
| Totaal | | | | | 214 338 |

Remarks on the Elaboration of an English–Spanish Word-combination Dictionary

Eduardo de Gregorio-Godeo, *Universidad de Castilla-La Mancha, Ciudad Real, Spain (eduardo.gregorio@uclm.es)*

Abstract: Through a focus on word-combination analysis in bilingual dictionaries, this article presents the epistemological and methodological background for an ongoing research and editorial project aiming to produce an English–Spanish dictionary of multi-word combinations. After a discussion of the treatment of word combinations as a phenomenon in different types of dictionaries, the lexicographic process guiding the elaboration of the dictionary is described. In addition to the preliminary plan and the fieldwork undertaken for this dictionary-making project, a detailed account is given of the principles determining the inclusion of entries and their presentation within the dictionary. The typology of word combinations included in the dictionary and the structure of entries is accordingly dealt with prior to making some remarks regarding the final presentation-and-revision stage currently being carried out. Because this article is based on current research intending to compile an English–Spanish dictionary of multi-word expressions, emphasis is constantly laid on a usage context where the source language is English and the target language Spanish. By considering the potential users of the English–Spanish word-combination dictionary examined here, some concluding remarks are made with regard to the educational implications of this kind of dictionary primarily aimed at intermediate- to advanced-level Spanish-speaking EFL learners.

Keywords: MULTI-WORD EXPRESSIONS, MULTI-WORD COMBINATIONS, COLLOCATIONS, IDIOMATIC EXPRESSIONS, ENGLISH–SPANISH BILINGUAL DICTIONARIES, LEXICOGRAPHY, PHRASEOLOGICAL UNITS

Opsomming: Opmerkings oor die bewerking van 'n Engels–Spaanse woordeboek van woordverbindings. Deur op die ontleding van woordverbindings in tweetalige woordeboeke te fokus, gee hierdie artikel die epistemologiese en metodologiese agtergrond vir 'n voortgesette navorsings- en redigeringsprojek met as doel die totstandbrenging van 'n Engels–Spaanse woordeboek van meerwoordige verbindinge. Na 'n bespreking van die behandeling van woordverbindings as 'n verskynsel in verskillende woordeboeksoorte, word die leksikografiese proses beskryf wat die bewerking van die woordeboek rig. Benewens die voorlopige plan en die veldwerk gedoen vir hierdie woordeboeksamestellingsprojek, word 'n uitvoerige verslag gegee van die beginsels wat die insluiting van inskrywings en hul aanbieding in die woordeboek bepaal. Die tipologie van die woordverbindings ingesluit in die woordeboek en die struktuur van die inskrywings word dus ook behandel voor daar 'n aantal opmerkings gemaak word oor die finale aanbieding- en hersieningstadium wat tans voltooi word. Omdat hierdie artikel gebaseer is op huidige navorsing met die doel om 'n Engels–Spaanse woordeboek van meerwoordige uitdruk-

kings saam te stel, word die klem gedurig geplaas op 'n gebruikerskonteks waar die brontaal Engels en die doeltaal Spaans is. Deur inagneming van die potensiële gebruikers van die Engels-Spaanse woordeboek van woordverbindings wat hier bespreek word, word 'n aantal slot-opmerkings gemaak met betrekking tot die opvoedkundige implikasies van hierdie soort woordeboek wat primêr gerig is op middelbare- tot gevorderdevlak- Spaanssprekende EVT-aanleerders.

Sleutelwoorde: MEERWOORDIGE UITDRUKKINGS, MEERWOORDIGE VERBINDINGS, KOLLOKASIES, IDIOMATIESE UITDRUKKINGS, ENGELS-SPAANSE TWEETALIGE WOORDEBOEKE, LEKSIKOGRAFIE, FRASEOLOGIESE EENHEDE

1. Introduction

Taking word-combination analysis in bilingual dictionaries as a case in point, this article presents the epistemological and methodological background underlying a research and editorial project aiming to produce an English-Spanish dictionary of word combinations. After a discussion of the treatment of word combinations as a phenomenon in different types of dictionaries, the lexicographic process guiding the elaboration of the dictionary is described. An account is accordingly made of the principles determining the inclusion of entries and their presentation within the dictionary. Given that this article is based on current research aiming to build an English-Spanish dictionary of multi-word expressions, emphasis will be laid at all times on a usage context where the source language is English and the target language is Spanish.¹ After the potential users of the English-Spanish word-combination dictionary here presented have been considered, some final remarks are made about the educational implications of such a dictionary primarily aimed at intermediate- to advanced-level Spanish-speaking EFL learners.

2. Word combinations and lexicographic analyses: a focus on collocations and idioms

Within the domain of lexicology, phraseology is progressively passing through a certain process of segregation as a separate branch of linguistics. While lexicology has to do with words and their meanings, phraseology focuses on such collocations of words (i.e. 'phraseologisms', 'phraseological units', 'collocations' or 'idioms'), where the meaning of the whole is different from the sum of the literal meanings of the words, comprising a phraseological unit. 'Phraseological units', which are the object of study of phraseology, happen to be stable word-groups with partially or fully transferred meanings. According to Gläser (1998: 125), a 'phraseological unit' may be conceived of as a lexicalized, reproducible, billexemic or polylexemic word group in common use, which has relative syntactic and semantic stability, may be idiomatized, may carry connotations, and may have an emphatic or intensifying function in a text.

As major types of phraseological units, 'collocation' and 'idiom' are familiar terms for users of English monolingual and bilingual dictionaries having achieved a certain linguistic proficiency. Such linguistic phenomena are often conceived of as 'multi-word combinations' (Ilson 2002: 333) whose meaning is more than the sum of the meaning of their components. As Benson, Benson and Ilson maintain in their approach to this lexical phenomenon, 'word combinations' come into being when "certain words regularly combine with certain other words or grammatical constructions" (1997: ix). To a great extent, when exploring the word combinations of a language, both collocations and idiomatic expressions are at some point examined.²

The notion of 'collocation' refers "to the tendency for certain words to occur together. The term itself comes from the verb *collocate*, meaning 'to go together'" (Finch 2000: 152).³ It is common practice for monolingual dictionaries to include information about units 'above' the word level, such as "units including more than one complete word, i.e. compounds and idioms like *black-bird, bank on, give up, night owl, hammer and tongs, at all, kick the bucket*" (Ilson 2002: 333). As substantiated by Ilson's investigation of lexicographic practices, most dictionaries generally incorporate syntagmatic information about the use of items in forming sentences, which may include notes on complementation and, which is the case in point here, information on "collocation with specific words or types of words ([e.g.] *fond of vs fondness for; [or] the association of capsize with boats or ships*)" (2002: 335).

According to Bussmann, collocations are characteristic word combinations "which have developed an *idiomatic* semantic relation based on their frequent co-occurrence" (1996: 81; emphasis added). Consequently, collocations are semantically and syntactically close to such lexico-grammatical units as 'idioms', that is, "a sequence of words which is semantically and often syntactically restricted, so that they function as a single unit. From a semantic viewpoint, the meanings of the individual words cannot be summed to produce the meaning of the idiomatic expression as a whole" (Crystal 2003: 225-226).

Idiomatic expressions are semantically related to collocations to such an extent that, as Crystal underlines, "an alternative terminology refers to idioms as 'habitual collocations'" (2003: 226). In fact, collocations and idioms may be taken to partake of the same stock of prefabricated units which, following the Russian tradition of phraseology, various authors have labelled 'word combinations' and also 'phraseological units' or 'phrasal lexemes' to delineate the "ready-made memorized combinations in written and spoken language" (Cowie 1998: 1), comprising both "'word-like' units, which function syntactically at or below the level of the simple sentence, and 'sentence-like' units, which function pragmatically as sayings, catchphrases, and conversational formulae" (Cowie, 1998: 4).⁴ Studies on the phraseology of English like that of Altenberg (1998: 120 *et passim*) likewise acknowledge idioms and collocations to be types of word-combinations. Indeed, Cowie himself highlights that "collocations of words in familiar literal senses are at one end of a broad spectrum

of word combinations in English. At the other are idioms: combinations whose constant re-use in a fixed form has led to a radical change of meaning" (1988: 131).

3. A project in context: English–Spanish word-combination lexicographic analysis

As Cowie asserted at the end of the last decade, there is now a "wider recognition of the crucial part that ready-made memorized word combinations "play in first- and second-language acquisition and adult language production [...] native-like proficiency in a language depends crucially on a stock of prefabricated units, 'prefabs', varying in complexity and internal stability (1998:1). Textbooks for EFL learners often examine collocations and idioms as fundamental aspects of vocabulary learning. This is also the trend in specific vocabulary-learning manuals for students of EFL (e.g. Thomas 1991; Redman 1997; McCarthy 2001, 2002). There are various monolingual dictionaries of idioms for EFL learners. A few examples of a long list that would be impossible to include exhaustively here are the *Longman Dictionary of Idioms* (Long 1979); the *Oxford Learner's Dictionary of English Idioms* (Warren 1994); and the *Collins COBUILD Dictionary of Idioms* (1995). Monolingual dictionaries of phrasal verbs as a special category of idiomatic expressions are also widely available on the market, for instance, the *Collins COBUILD Dictionary of Phrasal Verbs* (1989), including a workbook by Goodale (1993); the *Oxford Dictionary of Phrasal Verbs* (Cowie 1993) or the *Longman Phrasal Verbs Dictionary* (2000).⁵ There are likewise various monolingual dictionaries of collocations, including *Selected English Collocations* (Kozłowska 1993); *A Dictionary of English Collocations* (Kjellmer 1994); *The BBI Dictionary of English Word Combinations* (Benson, Benson and Ilson 1997); and the *Oxford Collocations Dictionary* (2004).

Current monolingual dictionaries for learners of English similarly incorporate information on collocations and idioms. Again, an exhaustive list could not possibly be included here, some well-known examples being the *Longman Language Activator* (1993); the *Longman Dictionary of Contemporary English* (1995); the *Oxford Advanced Learner's Dictionary* (Hornby 2000); and the *Collins COBUILD New Student's Dictionary* (2002). However, bilingual sources of English word combinations for Spanish learners of English in the form of user-friendly dictionaries are rather limited. Ordinary English–Spanish/Spanish–English dictionaries often include information on idioms and, to a lesser extent, collocations. Representative cases are the *Gran diccionario español–inglés, English–Spanish* (García-Pelayo y Gross 1993); the *Cambridge Word Selector, Inglés–Español* (1995); the *Diccionario Oxford avanzado para estudiantes de inglés, español–inglés inglés–español* (1996); *El Diccionario Oxford español–inglés, inglés–español* (2001); the *Diccionario de inglés contemporáneo para estudiantes: English–Spanish, español–inglés* (Sánchez Benedito and Gámez Gámez 2001); and the *Collins Dictionary: Español–Inglés, English–Spanish* (2003).

Although there exist English–Spanish dictionaries of phrasal verbs as a fundamental subcategory of idiomatic expressions like that of Khalaili and Marina (1984), comprehensive bilingual dictionaries of idiomatic expressions and collocations are almost non-existent. Thus, exceptions like the phraseological dictionaries of Carbonell Basset (1971, 1995, 1996) tend to lay a strong emphasis on proverbs and other idiomatic expressions, so that, in addition to disregarding a great deal of English idioms which are not proverbs, they do not on the whole take collocations into account. As substantiated by this overview of the literature on English–Spanish word-combination analysis, a thorough lexicographic examination of English–Spanish word combinations needs undertaking. Such a dictionary would be particularly useful for EFL learners whose mother language is Spanish. In fact, it is somewhat surprising that such a project should not have been contemplated to date, considering that word-combination dictionaries have already been produced which explore English phraseology from the viewpoint of languages with considerably fewer speakers than Spanish, for example, Polish (Osuchowska 2001).

4. Steps towards an English–Spanish dictionary of word combinations

Bearing in mind the above-mentioned lack of lexicographic resources providing translations of English word combinations into Spanish, we proceed to describe the approach followed for the elaboration of an English–Spanish dictionary of such phraseological units. Admittedly, this procedural account may serve as a basis for further lexicographic work exploring English multi-word combinations from the viewpoint of other target languages. The following subsections present the lexicographic process undertaken for the research and editing process leading to an English–Spanish dictionary of word combinations.

4.1 Lexicographic process

The project is consistent with the basic principles of contemporary lexicographic practices, thereby comprising all four stages of successful lexicographic work as mentioned by Hartmann (2001: 14-20), namely 'planning', 'fieldwork', 'description' and 'presentation.' The last stage includes final revision as well.

4.1.1 Preliminary plan

Hartmann stresses that "to be successful, lexicography as dictionary making requires careful planning and implementation of the compilation process on the basis of market research and the specification of the potential users' reference needs to be met" (2001: 20). As discussed above, an examination of the current availability of bilingual English–Spanish dictionaries gives evidence of

a significant lack of specific dictionaries dealing with word combinations. Therefore, a dictionary of this kind including collocations and idioms seemed to be a must, and, given the specificity of word combinations in every language, it seemed that the project would be beneficial for EFL learners as well as translators.⁶ After some preliminary work had been carried out through some local university- and regionally-funded research, the positive attitude of the Spanish Department of Education in granting funding on a national competitive basis confirmed the expected positive results of the project.⁷

4.1.2 Fieldwork

Further to the initial planning stage of the dictionary-making process, the first step in the research is to produce a thorough database of word combinations in English that might be used as a basis or 'macrostructure' for the alphabetically-ordered dictionary 'entries' in English, including Spanish equivalents, usage examples and other information within the English-Spanish dictionary of word combinations.⁸ The macrostructure of lexical entries has thus been completed on the basis of existing English monolingual dictionaries of multi-word expressions including *The BBI Dictionary of English Word Combinations* (Benson, Benson and Ilson 1997) or the *Oxford Collocations Dictionary* (2004), plus various other current English-Spanish dictionaries such as *El Diccionario Oxford español-inglés, inglés-español* (2001), and general-use monolingual English dictionaries like the *Oxford Advanced Learner's Dictionary* (Hornby 2000). Specialised dictionaries of English idioms, such as the *Collins COBUILD Dictionary of Idioms* (1995), have been consulted as well to build a macrostructure of word combinations in English consisting of collocations and idioms. Rather than come up with a new repertoire of English multi-word expressions defined in English, the main focus of the project was to provide Spanish equivalents for word-combination compilations in English, fieldwork therefore mainly relying on 'secondary sources'.⁹

The BNC (British National Corpus) (2001) has been adopted as a key source of information for the usage examples of the entries in the dictionary. Examples are easily retrieved through SARA, a PC-based concordance programme available on the commercial CD-based version of the BNC. The BNC has been chosen on grounds of its relevance for its present world-wide English-related lexicographic analyses: "the BNC has been used for the dictionaries of Oxford University Press, Longman and Chambers, the three publishers who contributed to its compilation" (Leech 2002: 91).¹⁰ However, when usage examples are absent in the BNC, other English-language corpora are consulted, for example, the *Collins COBUILD English Collocations on CD-ROM* (1995). The Internet is drawn upon as a last resource for usage contexts when no examples are found within existing language corpora.

Again, equivalents in Spanish for the word combinations in the macrostructure of the dictionary are provided on the basis of existing bilingual Eng-

lish–Spanish general-use dictionaries. A wide range of such bilingual dictionaries have been consulted in this regard. The list includes, but is not limited to, dictionaries like the *Diccionario Oxford avanzado para estudiantes de inglés, español–inglés inglés–español* (1996); the *Gran diccionario español–inglés, english–spanish* (García-Pelayo y Gross 1993); *El Diccionario Oxford español–inglés, inglés–español* (2001); the *Diccionario de inglés contemporáneo para estudiantes: English–Spanish, español–inglés* (Sánchez Benedito and Gámez Gámez 2001); the *Collins Dictionary: Español–Inglés, English–Spanish* (2003); or the *Cambridge Word Selector, Inglés–Español* (1995). When necessary, the CREA (Corpus de referencia del español actual, i.e. Reference Corpus of Contemporary Spanish) has been consistently used to test the appropriateness and idiomaticity of the equivalents in Spanish.

4.1.3 Description

4.1.3.1 Word combinations included in the dictionary

For the purposes of this dictionary, we have followed Benson, Benson and Ilson's (1986: 252-254) overall taxonomy of 'lexical combinations', namely: (a) 'free combinations' of words, which are those whose "components are the freest in regard to combining with other lexical items"; (b) 'idioms', which are "relatively frozen expressions whose meanings do not reflect the meanings of their component parts"; and (c) 'collocations', which are "fixed"/"recurrent" word combinations, that is, "loosely fixed combinations" between free word combinations and idioms.¹¹ Nonetheless, free combinations of words have not been considered in the dictionary, since their meaning may be found by examining the meaning of their individual constituent words in general dictionaries.¹²

4.1.3.1.1 Collocations

Benson, Benson and Ilson (1997) distinguish between 'grammatical collocations' and 'lexical collocations'.¹³ A 'grammatical collocation' is "a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause" (1997: xv); for instance, *decide on* (*decidirse por*) in *decide on a boat*. In contrast, 'lexical collocations' "do not contain prepositions, infinitives, or clauses. Typical lexical collocations consist of nouns, adjectives, verbs and adverbs" (1997: xxx); for example, *warmest regards* (*saludos afectuosos*) in *I send warmest regards*. The dictionary is mainly concerned with lexical collocations, which Spanish-speaking users of English might easily convert into wrong collocations (e.g. **deserted children/abandoned children* [*niños abandonados*]). Moreover, although there is a large number of English collocations following the Verb + Noun structure (*abandon hope*), this lexicographic repertoire does not include all possible free word combinations. In their discussion of English word combinations, Benson, Benson and Ilson (1997: xxx) make

this point clear by considering the construction *condemn murder* (*condenar delitos*), which is a free word combination in English, the verb *condemn* combining with an unlimited number of nouns (e.g. *condemn abortion, abuse of power*) and *murder* similarly combining with hundreds of verbs (e.g. *accept, acclaim, advocate*); however, *commit murder* (*cometer un crimen*) is a collocation. Following Benson, Benson and Ilson's (1997: xxx-xxxiii) taxonomy of collocations, the main types of lexical collocations included in the dictionary are the following:

(a) Verb + Noun / Pronoun / Prepositional Phrase

Most of these collocations consist of a verb of action or activation plus a noun or a pronoun (e.g. *come to a conclusion, make an appointment, set a record* [*llegar a una conclusión, concertar una cita, fijar un récord*]). In many cases, the collocation incorporates an idea of eradication (e.g. *override a veto* [*anular el veto*]).

(b) Adjective + Noun

Typical examples include *strong/weak tea* [*té fuerte/flojo*], which may not be replaced by *might/feeble tea*. Sometimes it is possible for various adjectives to collocate with a single noun. In these cases, only the most frequent lexical collocations are included. Furthermore, it is necessary to bear in mind that in English many nouns have an adjectival function when they are placed before another noun (e.g. *placement test* [*test de nivel*]). Such collocations appear in the dictionary entry of the second noun. Nevertheless, if the meaning of the second noun within the word combination is different from its meaning when used independently, the word combination is included in the dictionary on the basis of the second word (e.g. *acceptance speech* [*discurso de ingreso en una institución pública*]).

(c) Noun + Verb

These are collocations where the verb reflects characteristic actions of the noun, be it a person or a thing (e.g. *bombs explode/go off* [*las bombas explotan*]). Combinations which are easily predictable are not considered in the dictionary (e.g. *dancers dance, teachers teach, writers write* [*los bailarines bailan, los profesores enseñan, los escritores escriben*]).

(d) Lexical collocations indicating the 'unit' commonly associated with a noun

Typically, their structure in English is *noun₁ of noun₂*. Such collocations project the meaning that an individual belongs to a larger group (e.g. *a pack of wolves, a swarm of bees* [*una banda de ladrones, un enjambre de abejas*]), or the specific and concrete character of a unit with regard to larger elements (e.g. *a bit of advice, an act of violence* [*un consejo, un acto violento*]).

(e) Adverb + Adjective

Some examples of this type of collocation are *utterly abhorrent; patent-*

ly/totally/utterly absurd [*totalmente aborrecible; completamente, absolutamente absurdo*].

(f) Verb + Adverb

Here typical examples could be *appreciate deeply, greatly, keenly, sincerely, very much; argue bitterly, heatedly, passionately, strenuously, vehemently* [*agradecer sumamente, muchísimo; discutir apasionadamente, acaloradamente, vigorosamente*].

4.1.3.1.2 Idiomatic expressions

The dictionary takes account of a wide range of idiomatic expressions of contemporary British and American English. Idiomatic expressions fall within different categories such as: (a) traditional idioms (e.g. *spill the beans* [*irse de la lengua*]); (b) new phrases (e.g. *it's all gone pear-shaped* [*ha salido fatal*]); (c) metaphorical phrases (e.g. *face the music* [*afrontar las consecuencias*]); (d) two-word phrases (e.g. *wild card* [*comodín*]); and (e) various other similes (e.g. *like two peas in a pod* [*ser como dos gotas de agua*]).¹⁴ The dictionary does not incorporate 'phrasal verbs' as a characteristic type of multi-word combination because there are various dictionaries of phrasal verbs, both monolingual and bilingual, on the market.¹⁵

4.1.3.2 Structure of entries

The dictionary is articulated in accordance with the typical procedure of alphabetical order, which "is based on the written form of the lexically relevant units rather than on their meaning" and adopts a 'semasiological' approach in entries, that is, "going from name to notion" (Ilson 2002: 291). The basic principles guiding the organisation of entries within the dictionary are as follows:¹⁶

- (a) Key headwords, compounds included, are alphabetically ordered in the dictionary.¹⁷ Single-word compounds precede those written as two words. Homographs follow this order: adjective, adverb, noun, verb. For instance, *tease* (n.) precedes *tease* (v.). Determiners, prepositions and pronouns are not often headwords.

TEASE I n.

[person who teases] **a terrible tease** *un bromista*.

TEASE II v.

1. to tease a person cruelly *burlarse o reírse de una persona*. By then she had acquired a distinctive Geordie accent and she was upset when her friends at school teased her about her rounded vowels and up and down, sing-song voice.

2. to tease an animal *provocar a un animal*. He would meddle with fishing nets, pull up anchors and tow boats, tease dogs and tow swimmers.

3. to tease a fabric; wool *cardar un tejido; lana*. As a test of his strength Utnapishtim challenged him to stay awake for six days and seven nights —; But while Gilgamesh sat there resting on his haunches, a mist of sleep like soft wool teased from the fleece drifted over him (...).

- (b) Entries contain at least one context of usage. The key headword is written in small capital letters. English collocations are written in bold and Spanish equivalents appear in italics. Regular font face has been adopted for usage examples, the word combination in question being underlined. Here is an example:

MACHETE n.

to brandish, wield a machete *blandir un machete*. We can't take time to talk to the stylist before they wield the machete.

- (c) Cross-references within the dictionary are highlighted in yellow. For example, in the following entry *penny* is highlighted in yellow because the same idiom may be found in the entry for *penny* within the dictionary:

TEN n.

1. to be two/ ten a penny *ser baratísimo/no valer nada*. Uncritical testimonials to the postmodern's attractions are ten a penny, and conservative denunciations thereof not much scarcer.

2. ten to one (inform.) *diez a uno/te apuesto lo que quieras*. The great bulk of those who thronged Emmett place last night wanted to see the film, and they outnumbered protestors by a minimum of ten to one.

- (d) When a word collocates with others, the dictionary pinpoints series of collocations in an alphabetical order (e.g. **ABILITY: to demonstrate, display, exhibit, show**). This helps not only to save space, but also to display synonyms and near-synonyms. So, in the series for *ability*, *demonstrate ability* and *display ability* are treated as synonyms. However, collocations which are not synonyms are separated by a semicolon (;) (e.g. **domestic; physical; sexual ABUSE**). Synonyms are thus grouped together and separated by commas within the series of collocations. As shown in the example below, **to operate, run, use, work; shut down a machine**, *to operate, run, use* and *work* are treated as synonyms, and are likewise separated from *shut down* (which is not a synonym) by a semicolon:

MACHINE n.

1. to operate, run, use, work; shut down a machine *usar, utilizar; apagar una máquina*. A knowledge of how to set up and operate a machine could reduce the likelihood of a breakdown, so minimising the time spent on repairs.

- (e) Collocations whose meaning may be difficult to grasp by the student are often defined between brackets on the left of the collocation:

MARKET n.

[...]

5. **a bear ("falling"); bull ("rising") market** *mercado/bolsa a la baja; en alza*. Spicer points out that the sector has thrived in a bull market but the advertising boom may be coming to an end.

Information about register, dialects and other social factors is also indicated between brackets, e.g. formal (form.), American English (IAm), British English (IBr); etc:

MARINE n.

1. **a mercantile (IBr), merchant (IAm) marine** *marino mercante*. It is true that certain measures of nationalisation had been undertaken in the first months of Soviet government — for example, the Merchant Marine had been taken over in January 1918 (...).

- (f) As the following example of an entry from the dictionary instantiates, idiomatic expressions are written in blue at the end of entries:

MAT n.

1. **to weave a mat** *sacudir una alfombra/un felpudo*. The Lele are subsistence cultivators, growing maize, ground-nuts, and raffia palms from whose fronds they weave mats which are used as a special currency — a rare case of money really growing on trees.

2. **a bath; exercise; prayer; welcome mat** *alfombra de baño; para hacer ejercicio/colchoneta; para arrodillarse y rezar; de bienvenida*. For elderly people, grab rails at strategic places (including the bathroom/toilet) and use of a non-slip bath mat are devices likely to prevent accidents.

3. **to go to the mat (for sb)** (IAm) *hacer todo lo que puedes para solucionar un problema difícil, para ayudar a algn que tiene menos poder que tú*.

- (g) Examples have been taken from the BNC and, to a lesser extent, from the *Collins COBUILD English Collocations on CD-ROM* (1995). Although this is the case on very few occasions, the dictionary sometimes includes collocations without usage examples as no examples have been found in the language corpora utilized. However, an equivalent is always provided in Spanish, if the word combination is believed to be problematic for Spanish speakers:

AIR CONDITIONER n.

[...]

3. **to run; turn on an air conditioner; turn off an air conditioner** *poner/encender; apagar el aire acondicionado*.

- (h) Alternative translations of a collocation or idiom are separated by a slash (/):

MAD adj.

[...]

5. mad cow disease *enfermedad de las vacas locas/encefalopatía espongiiforme bovina*. Excuses abound: world markets have collapsed, diet-conscious Europeans are eating less red meat, some people in Britain fear it will give them mad cow disease.

- (i) Usage notes are highlighted in light blue, and range from pragmatic information to differences between British and American English, through other grammatical questions. They are also used to make reference to 'false friends', namely "terms in two languages which are phonologically and graphologically similar (cognates), but have subtly different meanings" (Malmkjær 2002: 82). Here is an example:

TAPE n.

[...]

9. audiotape; videotape *cinta de audio; de vídeo*. The screen replaces the car windscreen, and on it is shown a road scene that has been recorded on video tape, and transferred to disk.

Nota de uso: se pueden encontrar los compuestos juntos (videotape) o separados (video tape).

- (j) The use of a long underscore (____) in an idiomatic phrase indicates that various nouns, adjectives or verbs may be inserted in the gap:

MAN n.

[...]

33. our man in _____ *nuestro hombre en _____*. I don't know why our man in Madrid came to see me; just idle curiosity, perhaps — so few people have ever met me, so many seem anxious to do so.

4.1.4 Presentation and Revision

At this stage of the editing project, the dictionary on the whole is close to completion. The final stage of the project still needs to be undertaken. This will entail formatting, printing and proofreading the whole dictionary several times. A thorough revision of the dictionary will accordingly have to be carried out prior to its final publication.

5. Final remarks

Through the presentation of a research project attempting to produce an English–Spanish lexicographic repertoire of word-combinations using examples from the British National Corpus, this article has tried to shed light on the vital role of corpus-based language analysis for dictionary-making practices, chiefly as regards "bilingual dictionaries that have English as their source language"

(Benson 1985: 61). Focusing on collocations and idioms as fundamental multi-word-combination categories in English, this article has tried to disentangle the lexicographic mechanisms contributing to developing systematically-organised lexical repertoires of equivalents in Spanish.

This project is intended to compensate for the shortage of bilingual dictionaries providing instant and user-friendly access to Spanish translations of English word combinations. The fact is, as McCarthy stresses, that "the relationship of *collocation* is fundamental in the study of vocabulary" (1990: 12), and therefore is a major area of concern for learners of EFL. In a similar way, "idioms are a great source of difficulty for foreign learners" (Taylor 1990: 49), so that bilingual dictionaries of multi-word expressions are invaluable resources for both Spanish learners of EFL and translators. Dictionaries like this may thus help to facilitate and increase Spanish speakers' cultural awareness of English, since "all fluent and appropriate language use requires collocational knowledge" (Nation 2001: 318).

By way of conclusion, it could be said that an exercise like this may serve as a model for similar research and editing projects aiming to build word-combination bilingual dictionaries dealing with other languages. The production of a Spanish–English counterpart of this lexicographic repertoire is now the logical continuation of the present project.

Endnotes

1. According to Richards and Schmidt's definition of these terms when applied to bilingual dictionaries, the 'source language' is "the language out of which a translation is made" (2002: 496) and the 'target language' is "the language into which a translation is made" (2002: 539).
2. There is some degree of terminological confusion regarding word combinations and their subclasses. As Mel'čuk (1998: 23-24) evidences, the terms 'collocation' and 'idiom' — or 'idiomatic expression' — have sometimes been used to refer not only to a subclass of the word combinations of a language, but as synonyms of notions like 'word combination' or 'fixed (frozen) phrases'. His use of the term 'set phrases' (1998: 23ff), or 'phrasemes' as a synonym of word combinations to refer to phrases which are *not free*, is quite illuminating in this respect. Together with the term 'word combination', the lexicographic literature on phraseology recurrently draws upon the terms 'word-combination' (e.g. Howarth 1996), 'multi-word combination' (e.g. Ilson 2002) and 'multi-word expression' (e.g. Hartmann and James 1998) to refer to "a phrase consisting of two or more words functioning as a single lexeme. The constituents are relatively stable (fixed expression), and, if used idiomatically, their combined meaning is more or other than the sum of the parts, e.g. *fly-by-night*, *face the music*" (Hartmann and James 1998: 97).
3. The notion of 'collocation' was first introduced by Firth (1957) in his semantic theory to evoke the phenomenon of word combinations giving rise to semantic and idiomatic relations based on their co-occurrence.
4. See Cowie (1998) and Mel'čuk (1988) for detailed and complex terminological discussions outside the scope of this contribution.

5. Most of them are being constantly updated with new editions on the market.
6. This dictionary might accordingly be included within the category of dictionaries for the foreign learner, which are "essentially general-purpose dictionaries, but tailored to the needs of a specific group of users" (Jackson 1988: 174).
7. Preliminary contacts with a number of publishing firms similarly encouraged us to pursue such a project.
8. A dictionary 'macrostructure' is "a succession of articles, [that is,] entries, so ordered that any article may be found through an explicitly stable search procedure, an algorithm" (Ilson 2002: 291). An 'entry' is "the basic reference unit in a dictionary" (Hartmann and James 1998: 50). Entries consist of (a) a 'lemma' allowing compilers to locate and users to find entries in the word-list, and (b) a 'comment'. Comments may focus on 'formal' aspects (e.g. spelling, pronunciation, grammar) and semantic information (e.g. definition, usage, equivalents in the L2). As Hartmann and James add, "in case of multiple meanings of the lemma, the entry is subdivided into (usually numbered or otherwise marked) sections called 'sub-entries' or 'sub-senses', each of which provides the same basic information categories" (1998: 50).
9. During the fieldwork stage of dictionary making, 'primary' sources have to do with original material, whereas 'secondary' sources concern derivative material, chiefly other dictionaries (Svensén 1993).
10. As Summers states, "the British National Corpus is a collaborative initiative carried out by Oxford University Press, Longman, Chambers Harrap, Oxford University Computer Services, Lancaster University's Unit for Computer Research in the English Language (headed by Professor Geoffrey Leech), and the British Library. The project received funding from the UK Department of Trade and Industry and the Science and Engineering Research Council, and was supported by grants from the British Academy and the British Library" (1996: 266).
11. Benson, Benson and Ilson (1986: 252-254) also distinguish intermediate categories like 'transitional collocations', which fall between idioms and collocations, and 'compounds', that is, lexical elements of more than one word. Such categories are not contemplated here to avoid excessive complexity. The terms 'idiomatic expression' instead of 'idiom', and 'restricted collocations' instead of 'collocation', are sometimes found in the literature on word combinations.
12. A 'free combination of words' (Cowie 1981: 226) is a combination of words following only the general rules of syntax: the elements are not bound specifically to each other and so they occur with other lexical items freely; that is, the meanings of the words combine compositionally and may be substituted by synonyms (e.g. *run a business*, where *run* may be substituted by *manage*, in the same way as *a business* may be substituted by *a bank*).
13. The authors admit that terms like 'recurrent combination' or 'fixed combination' are sometimes employed instead of 'collocation' (1997: 15).
14. Cf. Long (1979: ix-x).
15. A 'phrasal verb' is "a unit in English which is formed from a verb with the addition of a preposition or adverb which can variously precede or follow an object: e.g. *take up*" (Matthews 1997: 279).
16. For further examples of entries in the dictionary, a final appendix has been included at the end of this article with the first entries of the letter *m* in the dictionary.
17. A 'headword' is the typographically canonical form of a word or a phrase which is chosen for the lemma, the position in the dictionary structure where the entry starts (Hartmann and James 1998: 68).

References

- Altenberg, B.** 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word-Combinations. Cowie, A.P. (Ed.). 1998: 101-122.
- Benson, M.** 1985. Collocations and Idioms. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 61-68. ELT Documents 120. Oxford: Pergamon Press.
- Benson, M., E. Benson and R. Ilson.** 1986. *Lexicographic Description of English*. Amsterdam/Philadelphia: John Benjamins.
- Benson, M., E. Benson and R. Ilson.** 1997. *The BBI Dictionary of English Word Combinations*. Second Edition. Amsterdam/Philadelphia: John Benjamins.
- BNC.** 1995–. *The British National Corpus*. Oxford: Oxford University Computing Services.
- Bussmann, H. (Ed.).** 1996. *Routledge Dictionary of Language and Linguistics*. London/New York: Routledge.
- Cambridge Word Selector, Inglés–Español: Diccionario Temático del Inglés Contemporáneo.* 1995. Cambridge: Cambridge University Press.
- Carbonell Basset, D.** 1971. *Diccionario de modismos inglés–español*. Madrid: Dos Continentes.
- Carbonell Basset, D.** 1995. *Diccionario fraseológico inglés–castellano, castellano–inglés, frases, expresiones, modismos, dichos, locuciones, idiomatismos, refranes, etc. A Phraseological Dictionary, English–Spanish, Spanish–English, Phrases, Expressions, Locutions, Clichés, Idioms, Sayings, Saws, Proverbs, etc.* Barcelona: Ediciones Serbal.
- Carbonell Basset, D.** 1996. *Dictionary of Proverbs: Sayings, Saws, Adages, English and Spanish — Diccionario de refranes: proverbios, adagios, castellano e inglés*. Barcelona: Ediciones del Serbal.
- Collins COBUILD Dictionary of Phrasal Verbs.* Sinclair, J. and R. Moon (Eds.). 1989. Glasgow: HarperCollins.
- Collins COBUILD Dictionary of Idioms.* 1995. Sinclair, J. and R. Moon (Eds.). Glasgow: HarperCollins.
- Collins COBUILD English Collocations on CD-ROM.* 1995. London: HarperCollins.
- Collins COBUILD New Student's Dictionary.* Sinclair, J. (Ed.). 2002. London: HarperCollins.
- Collins Dictionary: Español–Inglés, English–Spanish.* 2003. Barcelona: Grijalbo / London: HarperCollins.
- Corpus de referencia del español actual (CREA). Real Academia Española, Banco de datos [online]. <http://www.rae.es> [26 March 2010].
- Cowie, A.P.** 1981. The Treatment of Collocations and Idioms in Learners' Dictionaries. *Applied Linguistics* 2(3): 223-235.
- Cowie, A.P.** 1988. Stable and Creative Aspects of Vocabulary Use. Carter, R. and M. McCarthy (Eds.). 1988. *Vocabulary and Language Teaching*: 126-139. London/New York: Longman.
- Cowie, A.P.** 1993. *Oxford Dictionary of Phrasal Verbs*. Oxford: Oxford University Press.
- Cowie, A.P.** 1998. Introduction. Cowie, A.P. (Ed.). 1998: 1-20.
- Cowie, A.P. (Ed.).** 1998. *Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press.
- Crystal, D.** 2003. *A Dictionary of Linguistics and Phonetics*. Fifth Edition. Oxford: Blackwell.
- Diccionario Oxford avanzado para estudiantes de inglés, español–inglés inglés–español.* 1996. Oxford: Oxford University Press.
- El Diccionario Oxford español–inglés, inglés–español — The Oxford Spanish Dictionary Spanish–English, English–Spanish.* 2001. Oxford: Oxford University Press.
- Finch, G.** 2000. *Linguistic Terms and Concepts*. London: Macmillan.

- Firth, J.R.** 1957. Modes of Meaning. Firth, J.R. (Ed.). *Papers on Linguistics, 1934–1951*: 190-215. London/New York: Oxford University Press.
- García-Pelayo y Gross, R.** 1993. *Gran diccionario español–inglés, English–Spanish*. Barcelona: Larousse.
- Gläser, R.** 1998. The Stylistic Potential of Phraseological Units in the Light of Genre Analysis. Cowie, A.P. (Ed.). 1998: 125-143.
- Goodale, M.** 1993. *Collins COBUILD Dictionary of Phrasal Verbs Workbook*. Glasgow: HarperCollins.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography*. Applied Linguistics in Action Series. Harlow: Longman.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hornby, A.S.** 2000. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- Howarth, P.A.** 1996. *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Lexicographica Series Maior 75. Tübingen: Max Niemeyer.
- Ilsø, R.F.** 2002. Lexicography. Malmkjær, K. and J.M. Anderson (Eds.). 2002: 331-339.
- Jackson, H.** 1988. *Words and their Meaning*. London: Longman.
- Khalaili, K. and B. Marina.** 1984. *English Phrasal Verbs in Spanish*. London: Edward Arnold.
- Kjellmer, G.** 1994. *A Dictionary of English Collocations*. Oxford: Clarendon Press.
- Kozłowska, C.D.** 1993. *Selected English Collocations*. Warsaw: Wydawn Nauk PWN.
- Leech, G.N.** 2002. Corpora. Malmkjær, K. and J.M. Anderson (Eds.). 2002: 84-93.
- Long, T.H. (Ed.).** 1979. *Longman Dictionary of Idioms*. Harlow: Longman.
- Longman Dictionary of Contemporary English*. Summers, D. (Ed.). 1995. London: Longman.
- Longman Dictionary of Phrasal Verbs*. Courtney, R. (Ed.). 2000. Harlow: Longman.
- Longman Language Activator: The World's First Production*. Summers, D. (Ed.). 1993. London: Longman.
- Malmkjær, K.** 2002. Contrastive Linguistics and Cross-Linguistic Studies. Malmkjær, K. and J.M. Anderson (Eds.). 2002: 82-84.
- Malmkjær, K. and J.M. Anderson (Eds.).** 2002. *The Linguistics Encyclopedia*. Second Edition. London/New York: Routledge.
- Matthews, P.H.** 1997. *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.
- McCarthy, M.** 1990. *Vocabulary*. Oxford: Oxford University Press.
- McCarthy, M.** 2001. *English Vocabulary in Use: Upper Intermediate*. Cambridge: Cambridge University Press.
- McCarthy, M.** 2002. *English Vocabulary in Use: Advanced*. Cambridge: Cambridge University Press.
- Mel'čuk, I.** 1988. Collocations and Lexical Functions. Cowie, A.P. (Ed.). 1988: 23-53.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Osuchowska, D.** 2001. *English at Work: An English–Polish Dictionary of Selective Collocations*. Warsaw: PWN.
- Oxford Collocations Dictionary: For Students of English*. Crowther, J. 2004. Oxford: Oxford University Press.
- Redman, S.** 1997. *English Vocabulary in Use: Pre-Intermediate and Intermediate*. Cambridge: Cambridge University Press.

- Richards, J.C. and R. Schmidt.** 2002. *Longman Dictionary of Language Teaching and Applied Linguistics*. Third Edition. Harlow: Longman.
- Sánchez Benedito, F. and F. Gámez Gámez.** 2001. *Diccionario de inglés contemporáneo para estudiantes: English–Spanish, español–inglés*. Madrid: Pearson.
- Summers, D.** 1996. Computer Lexicography: The Importance of Representativeness in Relation to Frequency. Thomas, J. and M. Short (Eds.). 1996. *Using Corpora for Language Research*: 260-266. Harlow: Longman.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Taylor, L.** 1990. *Teaching and Learning Vocabulary*. New York: Prentice Hall.
- Thomas, B.J.** 1991. *Advanced Vocabulary and Idiom*. London: Nelson.
- Warren, H. (Ed.).** 1994. *Oxford Learner's Dictionary of English Idioms*. Oxford: Oxford University Press.

Appendix: Examples of entries from the letter M in the dictionary

MA'AM

1. Wham bam thank you ma'am (oral) *encuentro sexual que no se vuelve a repetir*. He was, it appears, neither rapacious nor peremptory, no 'for kicks', no 'wham bam, thank you, mam!' — he tried to make sure both of them enjoyed the time; and he was fun.

2. wham bam *algo que ocurre de repente sin preaviso o preparación y que termina rápidamente*. The problem arises when you fall from a screen above and land on 'monkey features': wham bam, every single life lost! — smacked bums all round and general loss of brownie points.

MACE n.

[staff used as a symbol of authority] a **ceremonial mace** *maza ceremonial*. The Liberal Democrats complained that too many schools have outside lavatories; that the Tories were profligate in setting up a chair of maritime history at a local university; and that they had spent too much on a ceremonial mace.

MACHETE n.

to brandish, wield a machete *blandir un machete*. We can't take time to talk to the stylist before they wield the machete.

MACHINE n.

1. to operate, run, use, work; shut down a machine *usar, utilizar; apagar una máquina*. A knowledge of how to set up and operate a machine could reduce the likelihood of a breakdown so minimising the time spent on repairs.

2. an adding, calculating machine *una máquina calculadora*. In this social service, therefore, the institutional element consists not in the great organisation over which John Boyd-Carpenter presides, with its large and efficient staff and its famous calculating machine at Newcastle.

3. an answering; washing machine *un contestador; una lavadora*. They could not afford a washing machine.

4. a cash, money access (IAm); cigarette; slot (IBr), vending machine *un cajero; un máquina expendedora de cigarrillos; máquina expendedora de bebidas*. Film fans can put down 20fr in a slot machine outside the Cinema Museum for a current programme.

5. a composing, linotype, typesetting; copy, copying, duplicating; fax machine *linotipia; fotocopiadora; fax*. The fax machine started chuntering away.

6. a heart-lung; X-ray machine *máquina de circulación extracorpórea; de rayos X*. I'm sorry, sir, one of the guards replied in English, but we have to check your case with the X-ray machine.

7. an earth-moving; milking; milling; sanding; sewing machine *excavadora; ordeñadora; fresadora; pulidora; máquina de coser*. You can bring your own sewing machine and basic sewing equipment

8. a mincing machine (IBr; IAm meat grinder) *picadora de carne*. Miss Paula McCloskey, 26, also injured her other leg when she became entangled in a mincing machine at the M U P factory in Richill, Co Armagh, in June, 1990, the High Court in Belfast was told yesterday.

9. a fruit (IBr), **slot** (IAm); **pinball machine** (IBr pintable) *tragaperras; flipper millón*. There ain't no music playing or nothing like there usually is, just the sound of people talking and the clunking of the fruit machine.

Nota de uso: En IBr se encuentra a veces pintable en vez de pinball machine.

Revolutionizing Bantu Lexicography — A Zulu Case Study

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Belgium; Xhosa Department, University of the Western Cape, Bellville, South Africa; and TshwaneDJe HLT, Stellenbosch, South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: Zulu uses a conjunctive writing system, that is, a system whereby relatively short linguistic words are joined together to form long orthographic words with complex morphological structures. This has led to the so-called 'stem tradition' in dictionary making — for Zulu, as well as for most other Bantu languages. Given this lemmatization approach has been found to be inadequate for young learners (who fail to isolate stems), the development of a new approach was imperative for them, but until recently deemed impossible to implement. In this paper it is argued that it is now perfectly possible to reverse the unproductive trend, and to opt for the lemmatization of full words for all but one of the word classes in Bantu. This revolution is made possible thanks to the recent availability of relatively large corpora, with which the really frequent citation options may be pinpointed. Rather than a mission statement, this paper offers the result for all word classes. To do so, an actual guide to the use of a Zulu dictionary is re-represented and annotated.

Keywords: ZULU, BANTU, DICTIONARY, USAGE GUIDE, MINI-GRAMMAR, WORD CLASSES, STEM VS. WORD LEMMATIZATION, CORPUS, USER-FRIENDLY

Samenvatting: Bantoe lexicografie radicaal omgooien — een gevalsanalyse voor Zoeloe. Zoeloe maakt gebruik van een conjunctief schrijfsysteem, d.w.z. een systeem waarbij relatief korte linguïstische woorden vast aan elkaar geschreven worden met lange orthografische woorden tot gevolg, die ook nog complexe morfologische structuren vertonen. Dit heeft geleid tot wat men de 'stam traditie' in de lexicografie is gaan noemen — voor Zoeloe, alsook voor de meeste andere Bantootalen. Aangezien deze lemmatisatieaanpak ongeschikt is gebleken voor jonge gebruikers (die woordstammen maar niet kunnen isoleren), moest voor hen een nieuwe aanpak ontwikkeld worden. Tot voor kort werd zo'n aanpak echter als niet-implementeerbaar beschouwd. In dit artikel wordt geargumenteed dat het vandaag de dag perfect mogelijk is om de onproductieve trend om te keren, en om te kiezen voor het lemmatiseren van volledige woorden voor alle woordklassen op één na in Bantoe. Deze radicale ommezwaai werd mogelijk gemaakt dankzij het recent beschikbaar komen van relatief grote corpora, waarmee de echt frequente opties qua trefwoordkeuzes bepaald kunnen worden. In de plaats van louter de beschrijving van een doelstelling, biedt dit artikel oplossingen voor alle woordklassen. Daartoe wordt de effectieve gebruikersgids van een Zoeloe woordenboek voorgesteld en van commentaar voorzien.

Sleutelwoorden: ZOELOE, BANTOE, WOORDENBOEK, GEBRUIKERSGIDS, MINI-GRAMMATICA, WOORDKLASSEN, STAM- VS. WOORD-LEMMATISATIE, CORPUS, GEBRUIKERSVRIENDELIJK

1. The one-size-fits-all problem

Although dictionaries for the Bantu languages have been compiled for several centuries now, and although Bantu metalexigraphy is at least 150 years old (cf. Benson 1964), the field has been plagued throughout by what one could call the one-size-fits-all approach. By and large that 'size' has been to attempt to lemmatize all words from all word classes under their *stems*, no matter the target user envisaged. Arguably, from a strict morphological point of view, this is a perfectly valid and linguistically sound approach. Students at institutes of higher learning are confronted with it from day one of their studies, and end up mastering the system given their otherwise general grounding in linguistics and their year-long exposure to a dictionary culture in a variety of languages. Such students are typically non-Africans studying at universities in the West or East. When that same approach is used to compile dictionaries for elementary learners, who moreover have not had the chance to be exposed to any other dictionaries in their lives, the dictionaries have shown to be too challenging to use. Here the intended user is typically a mother-tongue speaker of a Bantu language, in need of a dictionary with local relevance.

The problem of the one-size-fits-all approach for Bantu lexicography has been noted before in the scientific literature, and it is often presented as a need to choose between stem lemmatization (more often than not seen as 'the right size') vs. word lemmatization (typically 'the wrong size', or at least looked down upon, especially by linguists). Half a century ago Benson, after surveying the field, concluded in favour of stem lemmatization as follows:

It is now right and proper to [...] make certain suggestions which could help future compilers of dictionaries of African languages, whoever they may be, to avoid some of the more obvious pitfalls. [...] there are no rules laid down for lexicographers, and whatever has been learnt by toil and sweat, by trial and error, is worth passing on. [...] One cardinal principle which emerges from our study is that everything which needs to be said about a stem or root should be channelled into one single full article, complete with citations if needed.

— Benson (1964: 78, 80, 82)

Merely a year later, however, in a discussion of a Luganda–English dictionary, Snoxall argued in favour of word lemmatization, as follows:

[E]ven many Baganda would have little idea under what root form they should look up many of the commonest words which they use. [...] The general principle of entering words in a dictionary under roots [...] could never be of great assistance [...] It would seem therefore that, although disappointing perhaps to etymologists, a decision to enter headwords in the form in which they are used in actual speech, as words possessing meaning, [...] will be welcomed by the great majority of the users of the dictionary.

— Snoxall (1965: 27-28)

The debate has raged on ever since, with Bennett writing two decades later:

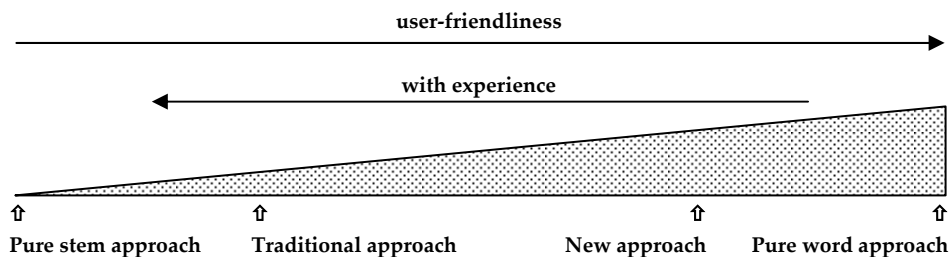
There has been debate as to the proper arrangement of the Bantu lexicon, and the question is far from settled. The inflection of nominals and verbals by means of prefixes, and the complex and productive derivational system, both characteristic of Bantu languages, pose difficulties [...] If items are alphabetized by prefix [...] a verb will be listed far from its nominal derivations, however transparent these may be. [...] A competing school arranges the lexicon by stem or root; this usefully groups related items, and saves on cross-referencing. Unfortunately, in such a system the user must be able to identify the stem, which given the sometimes complex morphophonemics of Bantu languages may not be easy.

— Bennett (1986: 3-4)

A more recent and excellent overview of the various pros and cons of the two opposing approaches can be found in Van Wyk (1995). The opposition, however, is a false opposition, as no Bantu dictionaries exist that are purely stem-based, neither do Bantu dictionaries exist that are purely word-based. I have pointed this out in an earlier contribution, one in which I had proposed a 'new approach':

These two extremes are but two poles on a continuum, of course. In reality, a 'traditional' stem-based approach to lemmatization [...] also has word features, and thus moves up on the continuum, while the approach advocated in this research article moves in the other direction of the continuum, away from the sole orthographic word. [The figure below] summarizes this situation, where the shaded triangle illustrates the increase in user-friendliness for *junior* users as one moves from stem-like to word-like lemmatization. With experience, however, one tends to crave for more condensed and more abstract information, and thus the wish to move in the other direction.

— De Schryver (2008a: 86-87)



It is my contention that the great majority of the existing dictionaries for the Bantu languages are actually variations of the 'traditional approach' depicted on the continuum, thus at heart dictionaries in which the lexicon is grouped around stems, with some dictionaries showing word-like features for selected word classes only, at which point they move up the continuum a little. Moving towards the word approach is easier to achieve for some Bantu languages than it is for others, depending on the degree of conjunctivism / disjunctivism of the

Bantu language in question (cf. Prinsloo and De Schryver 2002). Actually, no more than a handful of Bantu lexicographers have consciously tried to approach the continuum from the other (word) pole for *all* word classes. Moreover, and as we will see below, achieving this goal for the highly conjunctive Bantu languages has only become possible in recent years, thanks to the availability of (relatively) large corpora, from which real facts can be derived, rather than confining lexicographers to theoretical conjectures.

New types of dictionaries for the Bantu languages have indeed started to be compiled — that is, dictionaries for mother-tongue speakers, who do not have any previous exposure to a dictionary culture. Interestingly, the pressing need for such dictionaries was not only felt near-simultaneously in both monolingual and bilingual environments, the compilation of new, word-based dictionaries also started around the same time. For example, in Uganda, M. Nabirye undertook to compile the first monolingual dictionary for Lusoga, and in doing so she placed the (non-existent!) reference skills of the two million Basoga first. Her extensive research (Nabirye 2008) quite naturally led her to compile a dictionary in which *words* are listed, the only 'natural' language units according to the mother-tongue speakers she tested her dictionary material on (cf. Nabirye 2009a, 2009b). The result — the *Eiwanika ly'Olusoga* (Nabirye 2009) — is a most-powerful statement by a native speaker of a Bantu language.

Meanwhile in South Africa, several teams led by myself — a non-mother-tongue speaker of the Bantu languages — began work on a series of bilingual dictionaries with in each case one of the official South African Bantu languages and English as treated language pairs. The overarching theoretical framework was described in my PhD thesis (De Schryver 2004), and to date two dictionaries — the *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2007), and the *Oxford Bilingual School Dictionary: Zulu and English* (De Schryver 2010) — have been completed and published. An accompanying workbook was also prepared for the Northern Sotho dictionary (Taljard et al. 2008),¹ and both dictionaries have received favourable academic reviews; see for instance Prinsloo (2009) for the Northern Sotho dictionary, and Prinsloo (2010) for the Zulu dictionary.

Compared to any other earlier dictionaries for the Bantu languages at large, the *Oxford Bilingual School Dictionary: Zulu and English* (henceforth OZSD) is the first dictionary for a Bantu language to radically move away from the one-size-fits-all approach to lemmatization, in that the reference skills of the envisaged target user group forced me to come up with a new, tailored, word-based approach. In the present paper this approach is briefly introduced.

2. A daunting finding

Although the new approach is a direct implementation of the proposals which I formulated in my PhD thesis (completed in 2004), it would not have been possible to subsequently release two dictionaries at three-year intervals (for North-

ern Sotho in 2007, and for Zulu in 2010) without the support of large dictionary teams.² At this point I would like to salute and thank them, as *without* them my proposals would have remained academic theorizing; *with* them my dreams became a reality. For anyone wishing to make analogous dictionaries, it is instructive to have an idea about some of the practical aspects of these dictionary projects. I will use the OZSD as a case in point.

The OZSD is a bidirectional, bilingual school dictionary with Zulu and English as treated language pair, for use in South African schools in Grades 4 to 9 (i.e. the intermediate and senior phase of GET),³ covering 5 000 lemmas on each side. The OZSD has been designed to help students read and write Zulu better, if they are learners of Zulu, or read and write English better, if they are learners of English. The fact that it is a school dictionary had far-reaching implications for its design. Chief among those is that, in addition to the two A-to-Z sections, which amount to a total of 582 pages, the OZSD contains extensive extra-matter texts, totalling 58 pages (or ten percent of the A-to-Z sections), the main purpose of which is to teach a full-experience dictionary culture.⁴ Where relevant, those extra-matter pages are presented in both Zulu and English. As such, the front matter has a section on 'Dictionary features', in which the structure of the dictionary articles is visualised, and a gentle 'Introduction' on why dictionaries are important, on how the OZSD (a corpus-driven dictionary) differs from other dictionaries, and on what can be found in it. Completing the 12 pages of the front matter are also the title and imprint page, and a table of contents.

The middle matter or 'Study section', placed between the Zulu to English and English to Zulu sides of the dictionary, opens with 'Dictionary activities' in Zulu and English, taking the learners all the way from the sequence of the alphabet to both decoding and encoding exercises with the dictionary. Writing skills are covered in samples of e-mails, letters and electronic (SMS) messages. For Zulu, a mini-grammar and information on pronunciation are included, and for English, irregular verbs, punctuation and spelling are covered. All these sections, together with a detailed table of contents, amount to 30 pages.

The back matter has been conceived as a 'Reference section', and opens with six full-page plates with illustrations (domestic, wild and sea animals; small creatures; fruit and vegetables; etc.), and further contains information on the South African provinces, languages and phases of education, as well as numerous tables that bring together closed-class items (months, days, seasons; the solar system; public holidays; symbols; etc.), the numbering system in Zulu, and weights and measurements. Together with a detailed table of contents, as well as a page with the answers to the dictionary exercises from the middle matter, this back matter contains 16 pages.

Coordinating all these different components of the OZSD is no small matter, and considerable pressure was put on the human as well as financial resources — there never seemed to be enough of either of them. Although a tiny pilot was run in 2005 already (cf. De Schryver 2006), and although a preliminary draft of the Zulu to English side had been compiled by 2006, it is only

with a third team that the finish line was reached. In addition to myself as the editor-in-chief, the final dictionary development team consisted of a chief compiler (Nomusa Sibiyi) and a linguist (Arnett Wilkes), six more compilers (Sibusiso Dlamini, Thandeka Cebekhulu, Wo Mthembu, Mduduzi Ndlovu, Moses Biyela, Kholiswa Sitole), two proofreaders for Zulu (Msawakhe Hlengwa, Thokozani Buthelezi), two proofreaders for English (John Linnegar, Celia Slater), a consultant on curriculum entries (Daphne Paizee), and two computational engineers (David Joffe, Malcolm MacLeod). At the publishing house, a publishing manager (Megan Hall), two project managers (Fred Pheiffer, Phillip Louw), an editorial assistant (Lorna Hiles), three designers (Peter Burgess: A-to-Z, Oswald Kurten: extra matter, Sharna Sammy: cover), and two illustrators (Julien Marais, Leigh-Anne Wolfaardt) supported and interacted with the dictionary development team. Finally, two typesetters (Tommy Bell: A-to-Z, Ingrid Richards: extra matter) took care of the final layout. That's a total of 27 people to prepare the manuscript!

I have chosen to list all these team members here, rather than hidden in an endnote, as I want to make sure that it is clear from the start that everything possible was done to ensure that all the best available skills were brought together to create the OZSD. While the team at the publishing house was largely responsible for the development of an English dictionary-template for the English to Zulu side, to be translated into Zulu and coordinated with the material on the Zulu to English side by the dictionary development team, the main task of the dictionary development team itself was the creation of the Zulu to English side from scratch. To do so, that team had a relatively large Zulu corpus at their disposal, with which the Zulu lemma list could first be drawn up, and then queried for each and every lemma. The various meanings for each lemma were mapped directly onto the uses as seen in the corpus, and each main meaning was illustrated with material taken straight from the corpus.

All of this, of course, is easier said (or 'recounted' here) than done, and to do justice to all the details of the processes followed in compiling the OZSD, one would need far more pages than those that are available in a single scientific paper. There is firstly a big difference between the ways in which the two A-to-Z sides of the dictionary were compiled, an aspect with implications for the (non-)reversibility of the dictionary. That would have to be explained in a paper. Detailing how the Zulu lemma list was created would need at least one other paper-length treatment. A further paper could then deal with the description of the overall macrostructural as well as microstructural decisions. More detailed studies for the English side would have to focus on the presentation of the Zulu lexicon in the microstructure, as one can assume that general lemmatization decisions need not be covered for English. Grouping certain word classes or parts of speech (POSs), this could be done in about five papers. Detailing the lemmatization decisions as well as the dictionary structure on the Zulu side, however, can only be done in earnest if each and every word class is considered in isolation first, with generalizations in a second phase. There are

21 main Zulu word classes in the OZSD, and 42 if one includes the sub-word classes. Attention should furthermore also go to all the extra matter texts, conceived as front, middle and back matter in the OZSD. Together, I reckon about ten papers would be needed for a proper coverage of the various extra matter features. In short, then, giving a full account of the many aspects revolving around the creation of a dictionary such as the OZSD would require anything between 40 to 60 scientific papers. At about 20 pages each, this amounts to 800 to 1 200 pages in total, the equivalent of three full-length monographs!

This is a daunting finding, and one that has worried me for some years now. It is all good and well to write overview papers, in which one briefly sketches the general approach (cf. e.g. De Schryver 2008), but if one also wishes to stimulate a healthy academic debate, then more detailed studies are required, studies in which each step of one's reasoning is carefully argued. Specifically for the OZSD, I have presented detailed accounts for four Zulu word classes to date: possessive pronouns (De Schryver and Wilkes 2008), adjectives (De Schryver 2008a), quantitative pronouns (De Schryver 2008b), and ideophones (De Schryver 2009). Rather than to continue with the series of Zulu word classes, which at the current rate will take at least another decade, and rather than to give yet another overview (this time for Zulu, rather than for Cilubà, Swahili and Northern Sotho, as in De Schryver 2008), I have opted for a compromise in this paper. The OZSD itself actually contains a text that is particularly fit for this purpose, and this text is presented next.

3. How to use your dictionary (a Zulu mini-grammar)

On pages S13 to S26 of the Study section, thus right in the middle of the dictionary, the OZSD contains a chapter titled 'How to use your dictionary (a Zulu mini-grammar)', which is both a mini-grammar of Zulu in disguise and a true guide to the proper use of the Zulu dictionary. In the fourteen sub-sections that follow, I will present the text of that chapter in full, and I will intersperse it with additional comments. In doing so I hope to achieve at least four goals. Firstly, by presenting the full text of one of the extended extra-matter texts of a published dictionary, I illustrate the features of an actual *text in action*, rather than a proposal which may or may not materialize. Secondly, by using this particular text, I will automatically cover a *full system*, and run through all parts of speech of a language, thus covering the breadth one expects to find in overview papers. Thirdly, in the process the reader will have been offered a *synoptic view* of the grammar of Zulu relevant for lexicographic purposes, an aspect especially welcome to those not familiar with this Bantu language. And fourthly, with the added comments this paper functions as a stand-alone text, in which just enough depth is presented to serve as a *launching pad* for further academic discussions.

This approach remains an experiment, however, as two writing styles will now alternate throughout Section 3. On the one hand a dictionary user is

addressed, who is a Grade 4 to 9 learner, and for whom everything that is said is assumed to be new. I address that user directly, and avoid, wherever possible, all unnecessary 'difficult words'. On the other hand I am addressing the reader of the present paper, in an academic register, for whom using the correct terminology is crucially important, and for whom enough context and references must frame the work. The presentation starts with the text as found in the OZSD, and the added material is flagged with numbered comment fields in a smaller font size.⁵

3.1 Introduction

In this dictionary English words (in the English to Zulu side) have been listed as in any other English dictionary. If you are uncertain how to use a dictionary and would like some practice, please work through the 'Dictionary activities' first (see pages S4 to S6).

Comments 1: From the start a clear division is made between looking up lemmas in the English side vs. looking up lemmas in the Zulu side of the OZSD. The 'Dictionary activities' the user is referred to have also been prepared in Zulu (to be found on pages S2 to S4). Note the use of the terms 'word' and 'listed' rather than 'lemma' and 'lemmatized'. Following these two opening sentences, the remainder of the mini-grammar deals exclusively with the lemmatization and treatment of the Zulu lexicon, as seen below.

Zulu words (in the Zulu to English side) have been listed in a radically new way, unlike the approach used in any other dictionary for Zulu. In this dictionary you can look up many Zulu words *directly* as they are written and used. This is different to conventional Zulu dictionaries, where you need to break down most words until you reach the roots or stems of the words — and it is the roots or stems which have been listed in those dictionaries. In contrast, in this dictionary all primary prefixes are still attached to the word roots and stems, except in the case of verbs. For verbs you will still need to learn to cut off all verbal prefixes, as well as some other formatives. Therefore, studying and knowing Zulu grammar remains very important. You can only use this Zulu dictionary with success if you use it *together* with your Zulu textbook. What we have done in this dictionary is to make it much easier for you to find the words you need.

Comments 2: This section introduces the main revolutionary aspect of the OZSD, viz. the fact that for all Zulu word classes except one, an orthographic word was chosen as lemma rather than a root or stem.⁶ No information is given in the Zulu mini-grammar as to *how* words were chosen, but the use of a corpus is mentioned in the Introduction which is found in the front matter of the dictionary. The OZSD is not the place to explain this, nor is the present paper (as a paper each is needed per word class, cf. the penultimate paragraph of Section 2 above). What the Zulu mini-grammar does do is to explain the *result* for each word class. I can add here that for each word class as a whole, overall corpus frequencies were used in deciding where to cut off the various morphemes for that word class, so as to arrive at the most frequent type of orthographic form to be used as dictionary citation form for the entire paradigm. Making informed decisions to achieve this would

not have been possible without the availability of a relatively large Zulu corpus, consisting of a 7.5-million-word general-language component and a 1-million-word customised component of school textbooks.⁷ Throughout the scientific literature (cf. e.g. the references to earlier work in Van Wyk 1995), lemmatizing a conjunctively-written language such as Zulu as words rather than stems, has even been considered a theoretical impossibility due to the multiplication effect. With enough evidence of real use, however, and especially with frequency distributions for all words in all word classes — with which one is able to separate the wheat from the chaff, or thus the truly common (and frequent) from the rare (and infrequent) — it becomes possible to select exactly those (frequent) Zulu *words* that ought to feature in a dictionary aimed at junior learners, rather than having to resort to lemmatizing *roots and stems* only, for which only generic meanings rather than customised meanings and customised examples can be provided. The prime word class for which the word vs. stem debate has raged is that of the nouns, but it is as relevant for nearly all other word classes as well (adjectives, relatives, (derived) adverbs, pronouns, etc.). Not having the space for further elaborations here, I would like to refer the reader to De Schryver (2008a), where, as a case study for one of the word classes traditionally lemmatized as stems only, the treatment of adjectives as words is presented. For the overall procedure to arrive at the corpus-driven selection of all 5 000 Zulu lemmas, see especially page 69 therein.

Not surprisingly, 'the most frequent type of orthographic form' for a particular word class often corresponds with what is intuitively 'the most logical', with the outcome that each resulting lemma is also felt to be 'the most natural word'. As straightforward as this may seem, there were still numerous additional decisions that had to be taken, and many ad-hoc solutions had to be designed due to the fact that language is not as regular as linguists would like it (or force it) to be. As a result, the Zulu lemma list is not only unique in that it presents words rather than stems, but it is also unique to this very dictionary, as it is unlikely that another team of compilers for a Zulu dictionary would arrive at the very same corpus-driven lemmatized frequency list. The dictionary development team felt so strongly about this particular selection of lemmas that a copyright was taken on it, as seen on the imprint page: "© Zulu text, including Zulu headword list, TshwaneDJe HLT 2010".⁸ Claiming a copyright on the lexicon of a language is of course ludicrous, but claiming a copyright on a tailored lemma-sign list is not.

Also note the insistence here (and repeated further down) on the fact that the OZSD and the mini-grammar itself need to be used in conjunction with the learners' Zulu textbooks. This should not be seen as a cop-out for the aspects not covered in the mini-grammar, but rather as admitting that a mini-grammar is just that: a *brief* overview in which one simply cannot treat everything.⁹

Here is an easy example of how to look up words in this dictionary. All the words in the following sentence *Phuza amanzi ngazo zonke izikhathi, ikakhulu uma kushisa* 'Drink water at all times, especially when it is hot' are shown below. For each word you need to look up, we have shown the word class and a translation:

| | | | | | |
|-------------------------|-----------------|---------------|---------------------|------------------|-----------------------|
| word in sentence: | <i>phuza</i> | <i>amanzi</i> | <i>ngazo</i> | <i>zonke</i> | <i>izikhathi</i> |
| entry in dictionary: | -phuza | amanzi | ngazo | zonke | izikhathi > isikhathi |
| word class + class no.: | verb | noun 6 | temp. adv. 8 | incl. q. pron. 8 | pl. noun 7/8 |
| translation: | drink | water | at | all | times |
| word in sentence: | <i>ikakhulu</i> | <i>uma</i> | <i>ku-</i> | | <i>-shisa</i> |
| entry in dictionary: | ikakhulu | uma | [not in dictionary] | | -shisa |
| word class + extension: | conjunction | conjunction | indefinite concord | | verb + causative |
| translation: | especially | when | it | | be hot |

As you can see, most of the words in this sentence can be found in the dictionary exactly as they appear in the sentence. The main exception is verbs, which need to be looked up under the first letter of their root or stem. To find the root or stem you need to cut off all verbal prefixes, if these are present. The first verb (*-phuza* 'drink') has no prefixes, so you can look it up directly. With the last verb (*kushisa* 'it is hot') you first need to cut off the indefinite concord (*ku-* 'it'), and then look up the verb stem (*-shisa* 'be hot'). Refer to your Zulu textbook for the correct use and meaning of all verbal prefixes, as well as for the sound changes that take place when attaching prefixes to one another and to roots and stems. In this dictionary, the most important verbal prefixes are listed in Tables 4 to 6, but remember that those tables do not replace the need for you to study and know the grammar of Zulu!

Comments 3: Although the example used to illustrate the dictionary system is a real example (and as a matter of fact, is one of the examples under the lemma *ngazo*²), it was chosen for its 'easiness' and the fact that the syntax in Zulu runs parallel to the English syntax. Implicitly, the difference with a pure stem approach is well illustrated, as there each of the orthographic words would need to be looked up under respectively *-phuza*, *-nzi*, *-zo(na)*, *-nke*, *-khathi*, etc. In some of the existing Zulu dictionaries, such as Doke and Vilakazi's *Zulu-English Dictionary* (1953²) or Dent and Nye-mbezi's *Scholar's Zulu Dictionary* (1995³), some of these indeed need to be looked up under their roots and stems, others have nonetheless been lemmatized under their full words, and for still others both options have been lemmatized.

Although a dictionary is not a grammar, all the important verbal prefixes have nonetheless been tabulated in the mini-grammar, again with the caveat that the learners should still consider their textbooks as well.

From this example, you can see that nouns have been listed as full words in your dictionary, so you will look up *amanzi* 'water' under the letter A in the alphabet (and *not* under N, the first letter of the noun stem, *-nzi*). Even (frequently used) plural nouns have been listed in your dictionary, so you can look up *izikhathi* 'times' directly. This is a unique feature of this dictionary, which makes it very easy to use.

Comments 4: After having briefly introduced the procedure to look up verbs, the next word class given a first brief attention is nouns. While verbs could be said to have been lemmatized like in traditional dictionaries (with this difference that extensive tables of verbal prefixes are included in the mini-grammar), the inclusion of full noun prefixes attached to noun stems is a radical departure from tradition (and also the main aspect singled out for review in Prinsloo 2010).

Verbs and nouns as two different word classes are used to introduce the concept of a 'word class' next.

So far we have said that you need to look up verbs under the first letter of their roots or stems, and that you need to look up nouns under their full forms. Verbs and nouns are two different **word classes**, so you can see that it is important that you know which word class a certain word belongs to. There are many different word classes in Zulu, and you will need to know how to look up words in each type of word class. We will explain this in detail below.

In addition to a word class, you will also often see a number or even a pair of numbers following the word class in your dictionary. This is because nouns

in Zulu are traditionally grouped in different pairs of **noun classes** — which have unique pairs of **noun class prefixes** — and many other words have to be ‘in harmony’ with those noun class prefixes. In the example sentence, ‘at all times’ is *ngazo zonke izikhathi* in Zulu. Because the class 8 plural noun *izikhathi* is used, the temporal adverb *ngazo* and the inclusive quantitative pronoun *zonke* also have class 8 forms. If the class 7 singular noun *isikhathi* ‘time’ had been used, then the phrase would become *ngaso sonke isikhathi* ‘all the time’, with all words now in class 7 forms. It is very important that you understand this **need for harmony** (the so-called ‘concordial agreement system’) in Zulu, because only then will you appreciate the reason for assigning both a word class and class numbers to words in your Zulu dictionary. The different concords that are prefixed to verbs also need to be in harmony with the classes of the nouns they refer to, which is why all tables with concords consist of many lines: one line for each class, and a different concord for each class. The numbering system itself has been agreed upon internationally, so it is good you learn and know it.

Comments 5: This paragraph summarizes the core of the Bantu concordial agreement system, which is linked to the classification of Bantu nouns in noun classes, and a Bantu-wide numbering system. Note how the concept is gently introduced, by referring (twice) to a need for harmony, before the proper linguistic description is used. From a lexicographic point of view, the need to indicate the word class for each lemma, *and* (where relevant) the need to also indicate the class number, is also explained. Needless to say, the indication of class numbers across the word classes is missing from all existing dictionaries for Zulu.

3.2 Word classes

Most of the words listed in your dictionary are nouns and verbs. Actually, as many as 45.3% of all the dictionary entries are **nouns**, and 15.5% are **verbs**. Unique to this dictionary is that certain **locative forms** derived from nouns have also been listed as headwords; they make up 12.4% of your dictionary.

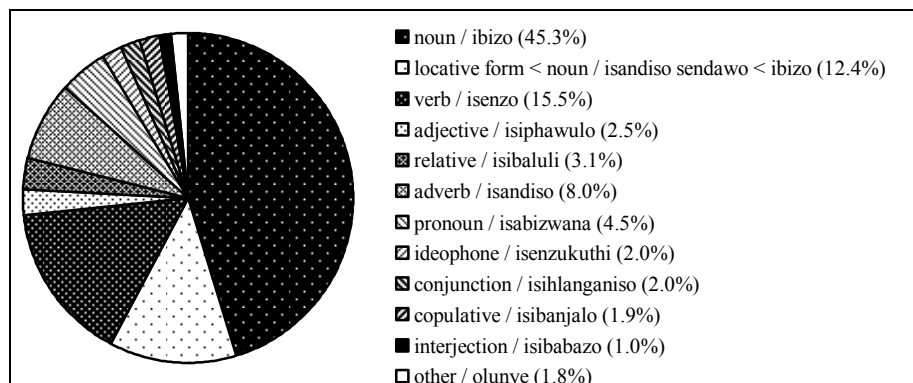


Figure 1: Distribution of the main Zulu word classes in your dictionary

As the pie chart shown in Figure 1 indicates, these three groups of words (nouns, verbs, and locative forms derived from nouns) make up nearly three-quarters of your dictionary. This does not mean that all the other word classes are less important — on the contrary. Many words in the other word classes are used much more frequently than some nouns and verbs. Words that ‘accompany’ nouns are typically **adjectives** and **relatives**; words that ‘accompany’ verbs are typically **adverbs**. To make well-formed sentences in Zulu, you also need to make use of **pronouns**, **ideophones**, **conjunctions**, **copulatives**, **interjections**, etc. We will now explain how you can look up words in each of these word classes.

Comments 6: With the concept of a word class introduced, it was now possible to extend the list of word classes to all the major ones in terms of number of members. In typical Zipfian style, some of the smaller word classes (in terms of members) actually contain the words used most often in Zulu (think for example of the class of the conjunctions). In the dictionary, actual word frequencies are indicated with a star rating following the lemma signs, an aspect explained in the Introduction to the OZSD. From the perspective of the dictionary user, the frequency breakdown shown in Figure 1 is most relevant, as it immediately tells that user how large each of the word classes is compared to the other word classes. The order shown in Figure 1 is also the order in which each of the word classes is given attention in the mini-grammar. That order combines both frequency (from most to least populous word class) and linguistic logic (e.g. bringing nouns together, or noun modifiers before verb modifiers, etc.). The coverage of each word class is henceforth fully driven by the relative frequencies seen in Figure 1. A word class with more members gets more attention than one with fewer ones, so on the whole this means that the sections become shorter and shorter as one proceeds through the different word classes in the mini-grammar. What is said about each word class is also fully driven by corpus facts, meaning that only what is most frequent in the corpus ends up being discussed. This is a radical departure from conventional grammatical descriptions, where one is interested in presenting full paradigms, irrespective of whether all items actually occur or not.

3.3 Nouns

Grammatically, a noun consists of a **noun class prefix** and a **noun stem**. The noun class prefix itself consists of a **pre-prefix** and a **basic prefix**. For example, the singular noun *isikole* ‘school’ consists of the pre-prefix *i-*, the basic prefix *-si-*, and the noun stem *-kole*. To change a singular noun into a plural noun in Zulu, you need to change the form of the noun class prefix, here to *izikole* ‘schools’. Nouns that have their singular in the *isi-* class and their plural in the *izi-* class belong to the class pair 7/8.

In your dictionary, you need to look up nouns under the pre-prefix; thus under the letter I for *isikole*. Under *isikole* you will find a full treatment of this word, with frequency of use and grammatical information, a translation, and an example sentence. The full treatment is normally only found under the **singular** form (*isikole*), and not under the plural form (*izikole*). When a plural form is frequent, it has also been listed in your dictionary, but the only information you will find there is a **cross-reference** to the singular form. So it is important

that you learn to recognize the full forms of nouns, with their pre-prefixes, and that you know how to change a plural noun into a singular noun. If you do not learn the system, you will spend more time thumbing through your dictionary, going from one entry to the next. The noun class system is shown in Table 1, which shows the main singular/plural noun pairs in your Zulu dictionary.

Table 1: Distribution of the singular and plural Zulu nouns in your dictionary

| Class (pair) | % | Class prefixes ① | Class prefixes ② | Notes |
|------------------------|--------------------|-------------------------|-------------------------|--|
| 1/2 1/2 | 4.6 3.7 | um-/aba- .. | umu-/aba- .. | ① Polysyllabic stem ② Monosyllabic stem .. |
| 1a/2a 1a/2a 1a/- | 4.6 2.1 1.8 | u-/o- .. u- | | |
| 3/4 3/4 3/- | 6.0 3.7 1.1 | um-/imi- .. um- | umu-/imi- .. umu- | ① Polysyllabic stem ② Monosyllabic stem |
| 5/6 5/6 5/- | 11.4 7.8 1.1 | i-/ama- .. i- | ili-/ama- | ② Variant for some monosyllabic stems |
| 6 | 1.1 | ama- | | |
| 7/8 7/8 7/- | 8.7 7.0 1.3 | isi-/izi- .. isi- | is-/iz- .. is- | ① Consonant-initial stem ② Vowel-initial stem |
| 9/10 9/10 9/- | 9.2 4.7 2.9 | iN-/iziN- .. in- | i-/izi- .. i- | ① With N = m before b, p, f and v; else N = n ② When nasal-initial stem, or with some loanwords .. |
| 11/10 11/10 | 2.7 1.4 | u-/iziN- .. | u-/izi- .. | ① With N = m before b, p, f and v; else N = n ② When stem starts with l or n |
| 11 | 1.6 | u- | ulw- | ① Consonant-initial stem ② Vowel-initial stem |
| 14 | 2.6 | ubu- | | + sound changes (=> <i>uboya</i> , <i>utshani</i> , <i>utshwala</i>) |
| 15 | 4.4 | uku- | ukw- | ① Consonant-initial stem ② Vowel-initial stem |
| Other | 4.7 | | | |
| | 100.0 | | | |

Table 1 tells you that most nouns in your dictionary — as many as 11.4% — belong to the singular class 5, with their corresponding plural in class 6. The notation 5/6 is used for these nouns: the number in bold (5) refers to the class number of the noun you are looking at, the other number (6) to the corresponding (here a plural) class number. Examples are *ikhanda* ‘head’, *ilanga* ‘day; sun’, *izwe* ‘country’. From Table 1 you can see that when you wish to look up the plural forms *amakhanda* ‘heads’, *amalanga* ‘days; suns’, *amazwe* ‘countries’, you can go directly to the letter I of your dictionary, to look up the singular forms.

For some class pairs, there are two sets of class prefixes. For example, for 1/2 one has both *um-/aba-* and *umu-/aba-*. The pair *um-/aba-* is used with noun stems that have more than one syllable (which we call 'polysyllabic stems'). For instance, the noun stem of *umfelokazi* 'widow' / *abafelokazi* 'widows' has four syllables: *-fe-lo-ka-zi*. The pair *umu-/aba-* is used with noun stems that have only one syllable (which we call 'monosyllabic stems'). For instance, the noun stem of *umukhwe* 'father-in-law' / *abakhwe* 'fathers-in-law' has only one syllable *-khwe*. Therefore, if you want to look up a noun that starts with *aba-*, Table 1 not only tells you that this is a plural noun, but also that the singular needs to be looked up under *um-* for a polysyllabic noun stem, and under *umu-* for a monosyllabic noun stem. For the class pair 3/4, one finds a similar situation: *um-/imi-* for polysyllabic stems, but *umu-/imi-* for monosyllabic stems.

Because nouns have been listed with their prefixes in your dictionary, you will not have problems looking up the alternative forms shown for the singular/plural pairs 5/6, 7/8, 9/10, and 11/10, nor for the alternative forms of class 11. As long as you remember to look up singular forms of nouns in full, you cannot go wrong. Similarly, 'difficult' class 14 nouns like *uboya*, *utshani*, or *utshwala*, can be looked up directly.

Comments 7: Table 1 is a quantified representation of the noun class system in Zulu, a first for this language. Noun classes are not considered in isolation, but are treated as genders, typically linking singular and plural members. In Column 1, dashes replace class numbers for one-class genders. Each number in bold, a notation first introduced in De Schryver (2001: 3-4), indicates the class of the member currently in focus. A noun class, as a member of a gender, is literally weighted (Column 2), and for each the various class prefixes are shown (Columns 3 and 4, with explanations in Column 5). The order of those class prefixes (with Column 3 always more frequent than Column 4) is based on dictionary occurrences, unlike the presentation in traditional grammars, where the so-called full forms are always presented first (e.g. *umu-aba-* before *um-aba-*). Only what is actually found in the dictionary is mentioned in Table 1 (e.g. for gender 9/- one finds the prefix *in-* rather than *iN-*, simply because no nouns in gender 9/- have been lemmatized that have the prefix *im-*, so there is no need to over-generalize). For linguists used to traditional Zulu grammars, a presentation like the one seen in Table 1 is undoubtedly a radical departure from more familiar presentations; for the user of the OZSD, however, this is simply in direct agreement with the word lemmatization used for nouns.¹⁰

Note, however, that when the pre-prefix of a noun is missing, as in *lo muntu* 'this person', you will first need to add the pre-prefix before looking up this noun under *umuntu* 'person'. Furthermore, various **morphemes** may be prefixed to a noun, and those need to be cut off. Thus, words like *lomuntu* 'of a person', *ngumuntu* 'she/he/it is a person', *nomuntu* 'with a person', *njengomuntu* 'like a person', *ngomuntu* 'about a person', *okomuntu* 'that of a person', *ngingumuntu* 'I am a person', *wayengumuntu* 'he was a person', etc. all need to be looked up under *umuntu*. Remember, therefore, that all nouns are listed under their pre-prefixes, which are either *a-*, *i-*, *o-*, or *u-*. Reformulated, all nouns have been listed under just four letters of the alphabet in your dictionary: A, I, O, or U.

Comments 8: Although lemmatizing nouns with their full prefixes, thus as complete words carrying meaning, is also the most intuitive lemmatization approach, users should not be led to think that all nouns appear in this canonical form in written Zulu. This paragraph dispels this, and shows typical environments of words and morphemes preceding nouns, either written disjunctively or conjunctively. Cutting off the conjunctively written parts may be challenging for the beginner, restoring the pre-prefix, which is always a reflection of the vowel of the basic prefix, should be more manageable.

In your dictionary you will also find over a hundred so-called **infinitive nouns**. These are all nouns derived from verbs, and always take the noun class prefix of class 15 (*uku-* or *ukw-*). These nouns have been chosen for a combination of two reasons: (i) they have new, independent meanings, and (ii) they are often used in Zulu. Examples are *ukuhlolwa* 'examination', *ukulimala* 'injury', or *ukwenza* 'action'. When the verb from which such an infinitive noun is derived has also been listed in your dictionary, you will see a cross-reference linking the noun to the verb, for instance *ukwenza* 'action' < *-enza* 'do, make, act; cause'. In general, all **deverbatives** (meaning all nouns which are derived from verbs) for which the verb is also listed in your dictionary, have been linked with that verb by means of a **cross-reference**. For example *impilo* 'life; health' < *-phila* 'live; be in good health'. Again, in this dictionary, you do not need to go to the alphabetic section PH (the first letters of the stem) in order to find the noun *impilo*; simply go directly to the first letter of the full noun, thus I.

Comments 9: The mini-grammar does not cover word formation processes, but cross-references in the dictionary do link deverbatives to their verb roots and stems whenever the latter have also been lemmatized. At the expense of fewer lemmas, cross-references in the other direction, thus from verb roots and stems to all the lemmas derived from those, could also have been considered. This was for example done in my *Cilubà-Dutch Lexicon* (De Schryver and Kabuta 1997) by means of what I termed '(frequency-based) tail slots' in my MA dissertation (De Schryver 1999: 53-54; cf. also De Schryver and Prinsloo 2001). Expecting Grade 4 to 9 learners to follow up on all cross-references away from a node, in addition to those from single spokes (deverbatives) to single nodes (verb roots or stems), was however considered too advanced for the OZSD.

When you do not find a certain noun under its singular form, this may be because the noun is infrequent and has thus not been listed in your dictionary. In some cases, however, the plural form was frequent enough to be listed, while the singular was not. In such cases, you will find a full treatment under the plural form. Examples include: *amaphesenti* 'percent; percentage', *iziphumuzi* 'punctuation marks', or *izinhlobonhlobo* 'different kinds'.

Lastly, note that not all possible singular/plural pairs have been listed in Table 1 (see the 4.7% 'Other'). Another combination is for example 9/6: *ifilimu* 'film, movie' / *amafilimu* 'films, movies', or *inkosi* 'chief, king' / *amakhosi* 'chiefs, kings'. Here too, your dictionary will show you the correct classes (9 and 6), as well as the exact forms (for example *inkosi* and *amakhosi*), so they should not be problematic to look up.

Comments 10: In line with what was noted under Comments 6, what is mentioned in the mini-grammar is restricted to what is both frequent and relevant for using the dictionary. Compared to Table 1, a staggering 19 further genders and combinations of genders are for example attested in the dictionary (nine of them hapaxes however), yet not knowing about those does not hamper successful dictionary use. Keeping them for a scientific article in which the noun class system of Zulu is revisited makes all the more sense. Likewise, there are also more (but infrequent) variant prefixes for some genders than the two columns of variants offered. For example, for class 11 the OZSD has one (and only one) lemma with the noun class prefix *ulu-*, namely *uluthi* 'stick; twig'.

3.4 Locative forms derived from nouns

Normally, a locative meaning can be 'added' to a noun from the class pairs 1/2 or 1a/2a by replacing the pre-prefix with the class 17 locative prefix *ku-* (or the variant *ko-*). As such, *umuntu* 'person' becomes *kumuntu* 'to a person', or *omalume* 'uncles' becomes *komalume* 'to the uncles'. Especially with loanwords in the class pair 9/10, one may also find the variant *kwi-*. As such, *inombolo* 'number' becomes *kwinombolo* 'at the number'. Word forms like these are not hard to decode, and have therefore not been listed in your dictionary.

A second way to add a locative meaning to a noun is to replace the pre-prefix with *e-* and to add the suffix *-ini*. As such, *abantu* 'people' becomes *abantwini* 'to/from/among/... the people', or *umthombo* 'fountain, spring' becomes *emthonjeni* 'in/at/to/from/... the fountain/spring'. This approach may be used for nouns in all classes, except for those in class 11 or 14, where the pre-prefix is replaced with *o-*, again with the suffix *-ini*. As such *uhlangothi* 'side' becomes *ohlangothini* 'on/at/to/from/... the side', or *ukhetho* 'election' becomes *okhethweni* 'in the election'. Numerous sound changes are found when the suffix *-ini* is attached to nouns, as seen in the underlined parts in the examples here. It is mainly for this reason that all frequent locative forms that are derived from nouns using the so-called *e-lo-...-ini* 'locativisation strategy' have been listed directly into your dictionary. Grammatically, these locativised nouns have actually become 'locative adverbs'. To save space in the dictionary, only translations into English of the meanings are given for these locative adverbs, without any examples.

A second reason for listing these locative adverbs in your dictionary is that there are locativised nouns where the suffix *-ini* does not appear. The appearance or not of the suffix *-ini* is not predictable, which makes it useful to list the correct frequent forms in the dictionary. As such, all of the following forms for example appear without the suffix *-ini*: *ebusika* 'in/during/... winter' (< *ubusika* 'winter'), *ekhaya* 'at/from/... home' (< *ikhaya* 'home'), or *olwandle* 'in/on/to/from/... the ocean/sea' (< *ulwandle* 'ocean, sea').

A third reason for listing these locative adverbs in your dictionary is that there are locativised nouns for which the frequency of the noun itself is extremely low (or may not even appear at all in our 8.5-million-word Zulu corpus). Examples include: *emaphandleni* 'in/to/from/... the rural areas', *emsamo* 'in/at/to/from/... the back of the hut', or *esidlangalaleni* 'in public; openly'. For

all these examples the corresponding noun is not seen in the corpus, so it would be wrong to list such nouns in a dictionary that is corpus-driven and that focuses on frequently used words.

Lastly, as was the case for the nouns, locativised nouns may be preceded by one or more **morphemes**, which need to be cut off before they can be looked up in your dictionary.

Comments 11: This sub-section is entirely self-contained and provides an excellent argued example of the type of choices that had to be made with regard to the lemmatization or not of certain Zulu 'words'. Looking up nouns locativised by means of *ku-* and its variants was considered manageable for the target user group, but locativisation by means of the *e-/o-...-ini* strategy was not. For a linguistic account of the issues involved, see De Schryver and Gauton (2002).

3.5 Verbs

In this dictionary verbs need to be looked up under the first letter of their roots or stems. The same approach is followed in all dictionaries for Zulu. This is because each verb root or stem can combine with very many combinations of prefixes, and the ending of a verb also varies (it can be *-a*, *-e*, *-i*, *-ile*, or *-anga*). Listing all these possibilities in a dictionary would not be practical, as one would need to list many dozens, sometimes hundreds of forms for each verb. This therefore means that you will need to learn to cut off all **verbal prefixes** and **formatives** from verbs (those listed in Tables 4 to 6 as well as all the others listed in your Zulu textbook) before you can look up the basic meaning of a verb.

Comments 12: Van Wyk (1995: 86-87) performed a quick, back-of-the-envelope calculation for Zulu, and claimed that "[t]he number of combinations possible for a suitable transitive verb stem is [...] $18 \times 19 \times 6 \times 2$," or thus 4 104. While the actual theoretical figure is many times higher (for one, Van Wyk did not take the variation in verb endings into account), corpus counts indicate that the number of those that are actually used is many times smaller. Disregarding the very rare uses, which would not have to feature in a word-based school dictionary anyway, the number of those that are used is still very high (many dozens to several hundred orthographic forms per verb root or stem, as stated in the mini-grammar), so a word-based approach to the lemmatization of verbs in a paper dictionary for Zulu is indeed not feasible.¹¹ This thus means that we have come back to square one as far as the lemmatization of verbs in a paper dictionary for Zulu is concerned, in that the traditional approach is stuck too. The exercise was not futile, however, as the decision was arrived at following a study of large amounts of actual language use, rather than being based on linguistic extrapolations. The decision, in other words, can truly be defended.

A verb without any verbal prefixes and without any verbal extensions is known as a **verb root**. In your dictionary the final vowel *-a* is always added to verb roots. Examples are *-anga* 'kiss', *-linga* 'try, attempt', and *-thuma* 'send'. 54.5% of all the verbs listed in your dictionary are verb roots. The other 45.5% of the verbs take one or more **verbal extensions**. When a verb takes a verbal extension, it is known as a **verb stem**. Verb stems are also shown together with the final vowel *-a* in your dictionary.

Comments 13: Some linguists may object to the terminology used here, and may prefer to talk about (formal or base) radicals and extended radicals rather than roots and stems respectively, and may prefer to view a stem as merely a root plus the verbal ending, with that root either a radical or extended radical (cf. e.g. Schadeberg 1992³: 8).

In Table 2 the distribution of the verbal extensions in your dictionary, as well as the main combinations of verbal extensions, is shown.

Table 2: Distribution of Zulu verbs with verbal extensions in your dictionary

| Verbal extension | Form | Meaning | % | Examples |
|------------------------|-------------------|----------------------------|-------|--|
| passive | -(i)w- | adds 'be', 'being' | 32.0 | <i>-akhiwa, -khishwa, -shaywa</i> |
| applicative | -el- | adds 'for', 'on behalf of' | 16.6 | <i>-buyela, -khulumela, -xoxela</i> |
| causative | -is- | adds 'cause to', 'help' | 13.6 | <i>-balisa, -bhalisa, -qondisa</i> |
| neuter | -ek- OR -akal- | adds '-able/-ible' | 12.1 | <i>-bhekeka, -bonakala, -tholakala</i> |
| reciprocal | -an- | adds 'each other' | 4.5 | <i>-bambana, -bhekana, -xhumana</i> |
| applicative + passive | -elw- | adds 'be ... for/by' | 4.5 | <i>-balelwa, -banjelwa, -phelelwa</i> |
| causative + passive | -isw- | adds 'be ...' | 3.9 | <i>-aziswa, -fundiswa, -setshenziswa</i> |
| causative + reciprocal | -isan- | adds '... each other' | 1.8 | <i>-bonisana, -hambisana, -phikisana</i> |
| reciprocal + causative | -anis- | adds '... each other' | 1.2 | <i>-ehlukanisa, -qondanisa, -xhumanisa</i> |
| perfective | -elel- | completes an action | 1.2 | <i>-azelela, -bambelela, -phumelela</i> |
| intensive | -isis- | expresses intensity | 1.2 | <i>-bhekisisa, -qondisisa, -zwisisa</i> |
| <i>Other</i> | | | 7.3 | |
| | | | 100.0 | |

Nearly one third (32.0%) of all the verbal extensions simply add a passive meaning to a verb root, 16.6% add an applicative meaning, 13.6% a causative meaning, etc. Because all these verb stems (thus 'verbs with verbal extensions') have been listed directly into your dictionary, you do not need to memorize the various sound changes that apply when suffixing certain verbal extensions to verb roots. For example: *-khipha* + passive > *-khishwa*. Also, the exact meanings of the resulting verb stems are mentioned in the dictionary articles of those verb stems, with examples to support those meanings. It is a good idea now to look up each of the examples listed in Table 2, and to compare the dictionary information for each verb stem with that found under each verb root from which it is derived. (Note that a **cross-reference** always links a verb stem with its verb root, on the condition that that verb root is frequent enough to be listed in your dictionary.)

Comments 14: Although I said that verbs in the OZSD are lemmatized like in any other dictionary for Zulu, this is not entirely true in that this is only correct for verb *roots* (or for formal or base radicals if one prefers). Existing dictionaries for Zulu will not systematically lemmatize verb *stems* (or extended radicals). Doke and Vilakazi (1953²), for example, do not include any verbs with passive extensions, and will furthermore normally only include those verbs with verbal extensions that have undergone some level of lexicalisation. In contrast, the OZSD systematically includes all frequent verb stems.

As an example of the exercises the dictionary user is asked to do, (1) and (2) below show the effect of adding a perfective extension:

- (1) **-phumelela** ** verb + perfective
C-PHUMA ■ succeed • Kufanele usebenze kanzima ukuze **uphumelele**. *You must work hard in order to succeed.*
- (2) **-phuma** *** verb ■ go out; come out; exit
• UGumede **uphume** emotweni ephethe izimbali ezinhle zikankosikazi wakhe. *Gumede got out of the car carrying beautiful flowers for his wife.*

This exercise will give you a good idea of the meanings added by the verbal extensions, and it will expose you to typical dictionary structures. Three of those structures are commented on below.

Firstly, note that frequent reflexive verbs are mentioned within the dictionary articles of the verbs from which they are derived, and thus need to be looked up there. To form a reflexive verb, the **reflexive concord** *-zi-* (or *-z-* before vowels) is simply prefixed to a verb root or stem. For example, for the applicative form *-xoxela* ‘tell (someone)’, which is derived from the verb root *-xoxa* ‘talk’, the reflexive form is *-zixoxela* ‘just talk, simply talk, merely talk’. This reflexive form is explicitly mentioned under the verb stem *-xoxela* as a derivation because most uses of that verb stem take the reflexive concord. Throughout your dictionary, all **derivations** are preceded by a black arrow (►). For examples of other derivations, see the verb *-bona* (> *sawubona*), the noun *imali* (> *malini*), or the adjective *enkulu* (> *enkulukazi*).

Comments 15: Two of the examples mentioned are shown in (3) and (4) below:

- (3) **-xoxela** verb + applicative C-xoxa ■ tell (someone) • Ungalokothi **umxoxele** uMaNdlovu ngaleli phupho lakho. *You dare not tell MaNdlovu about this dream of yours.*
► **-zixoxela** ■ just talk; simply talk; merely talk • Baqhubeka-ke **bezixoxela** abangane bethu laba bedlalisa nomntwana owayelokhu ehleka. *They continued merely talking, these friends of ours, while playing with the child who was laughing all the time.*
- (4) **imali** *** noun 9/10 (pl. **izimali**) ■ money
• Ukhokhe **imali** eningi yesikole. *She paid a lot of money in school fees.*
► **malini** ■ how much money? • Kukhokhwa **malini** emathekisini ahambela eMtuba esuka eMpangeni? *How much money is paid for taxis going from eMpangeni to Mtubatuba?*

Secondly, verbs with the reciprocal verbal extension (*-an-*) are mostly, sometimes exclusively, followed by the **adverbial formative** *na-* ‘with, together with’. For example, look up the verbs *-bhekana*, *-hambisana*, and *-phikisana*, and

you will see that translations and examples are only given for the combinations *-bhekana na-*, *-hambisana na-*, and *-phikisana na-*. Throughout your dictionary, all **combinations** are preceded by an empty diamond (◇). Also, in your dictionary, when verbs combine with nouns to form new meanings, you will need to look up the noun and not the verb for those meanings. See for example the combinations with verbs listed under the nouns *icala*¹, *ithambo*, or *intwala*.

Comments 16: Two of the examples mentioned are shown in (5) and (6) below:

- (5) **-phikisana** verb + causative + reciprocal
C-PHIKA
◇ **-phikisana na-** ■ argue (with); disagree (with); dispute; oppose • Musa **ukuphikisana** nam! *Don't argue with me!*
♦ Akekho **owaphikisana** nokuqanjwa kabusha kwalo mgwaqo. *There is no one who disputed the renaming of this road.*
- (6) **icala**¹ *** noun 5/6 (pl. **amacala**) 1 ■ case (in law) • Kuyothi uma **icala** liya emajajini wena ube ngufakazi. *When the case goes to the judges, you will be the witness.*
2 ■ offence • Kuthiwa badedelwa kodwa baphinde benza amanye **amacala** anzima kakhulu kunalawo ababeshelwe wona. *It is said that they were released, but they committed other, more serious offences than those they had been arrested for.*
◇ **-bekwa icala** ■ be accused (of); be charged (with) • Waboshwa abekwe **icala** lokweqa umyalelo wenkantolo. *He was arrested and accused of breaking an order of the court.*
◇ **-vuma icala** ■ admit; plead guilty • Uma elivuma **icala**, uzomiswa imidlalo emithathu. *If he pleads guilty, he will be suspended for three games.*

Thirdly, **grammatical abbreviations** may also be used in derivations and combinations. These are used to summarize a full list (known as a 'paradigm') of possibilities. For example, if you look up the verb stem *-azelela*, you will find the derivation *[SC+]**ngazelele***. Here SC stands for subject concord, and this notation means that any of the subject concords seen in Table 4 must precede *-ngazelele* in order to obtain the meaning.

Comments 17: The article for the verb stem *-azelela* is shown in (7) below:

- (7) **-azelela** verb + perfective C-AZI¹
► **[SC+]**ngazelele**** ■ unexpectedly; surprisingly
♦ Wathi **engazelele** wabona ikhanda likaShaka livela ngaphezu kothango lwesigodlo. *He unexpectedly saw Shaka's head appearing above the fence of the royal enclosure.*

The various grammatical abbreviations used in your dictionary are listed in Table 3.

Table 3: Grammatical abbreviations used in your dictionary

| Code | Meaning | For examples, see under |
|--------|---|--|
| [DEM+] | preceded by any demonstrative | <i>lakho, lwethu, wethu</i> |
| [NEG+] | preceded by any negative | <i>elangeni, isinqe, quthu</i> |
| [PC+] | any possessive concord prefixed to word | <i>igolide, isihlanu, ngasese</i> |
| [RC+] | any relative concord prefixed to word | <i>-banda, -mela, -phakama</i> |
| [SC+] | any subject concord prefixed to word | <i>-azelela, ngokuphindwa, uzwelonke</i> |
| [+LOC] | followed by any locative | <i>-balekela, -sebenzela, -shonela</i> |

Comments 18: As before, the dictionary user is expected to look up one or more of the examples given, in order to get acquainted with the dictionary system. As an illustration, one of the examples for the grammatical abbreviation [DEM+] is shown in (8) below:

- (8) **wethu** ** possessive pronoun 1+1p pl ■ our
 ♦ Ngisho nomqeqeshi **wethu** unethemba elikhulu ngalo mdlalo. *Even our trainer has a lot of faith in this game.* 3+1p pl ■ our
 ♦ Kungumgomo **wethu** ukuphatha iziguli ngesizotha nangesineke. *It's our policy to treat patients with dignity and care.*
 ◇ [DEM+] ... **wethu** ■ of ours ♦ Kumele senzeni nje nempela kulo mjondolo **wethu**? *What must we really do in this shack of ours?*

Verbs are and will always be the most difficult part of speech to look up in a printed Zulu dictionary. This is so because a verb root or stem is normally 'hidden' in the middle of a much longer word. In order to help you recognize verb roots and stems, Table 5 lists the main verbal tenses and moods in Zulu. Table 5 has to be used together with Table 4, as well as Table 6. In a way, Tables 4, 5 and 6 give you a short overview of the verbal grammar. Please read through this information slowly, and reread it often, until you memorize it.

Table 4: Verbal prefixes (SC and OC) to be cut off before looking up Zulu verbs, versus adjective prefix (AP) and relative concord (RC) to keep when looking up adjectives and relatives

| Class no. | Class prefix | +SC +C | +SC +a/e | +SC +o | -SC | ~SC | §SC | OC | AP | RC |
|-----------|--------------|--------|----------|--------|-------|-------|-------|--------|--------|------|
| 1; 1a | um(u)-; u- | u- | w- | w- | ka- | e- | a- | -m(u)- | om(u)- | o- |
| 2; 2a | aba-; o- | ba- | b- | b- | = +SC | be- | = +SC | -ba- | aba- | aba- |
| 3 | um(u)- | u- | w- | w- | wu- | = +SC | = +SC | -wu- | om(u)- | o- |
| 4 | imi- | i- | y- | y- | yi- | = +SC | = +SC | -yi- | emi- | e- |
| 5 | i(li)- | li- | l- | l- | = +SC | = +SC | = +SC | -li- | eli- | eli- |
| 6 | ama- | a- | ∅ | ∅ | wa- | e- | = +SC | -wa- | ama- | a- |
| 7 | is(i)- | si- | s- | s- | = +SC | = +SC | = +SC | -si- | esi- | esi- |
| 8 | iz(i)- | zi- | z- | z- | = +SC | = +SC | = +SC | -zi- | eziN- | ezi- |
| 9 | i(N)- | i- | y- | y- | yi- | = +SC | = +SC | -yi- | eN- | e- |
| 10 | izi(N)- | zi- | z- | z- | = +SC | = +SC | = +SC | -zi- | eziN- | ezi- |
| 11 | u(lu)- | lu- | lw- | l- | = +SC | = +SC | = +SC | -lu- | olu- | olu- |
| 14 | ubu- | bu- | b- | b- | = +SC | = +SC | = +SC | -bu- | obu- | obu- |

| | | | | | | | | | | |
|-------|------|------|-----|-----|-------|-------|-------|-------|-----------|-------|
| 15 | uku- | ku- | kw- | k- | = +SC | = +SC | = +SC | -ku- | oku- | oku- |
| 17 | ku- | ku- | kw- | k- | = +SC | = +SC | = +SC | -ku- | oku- | oku- |
| 1p sg | — | ngi- | ng- | ng- | = +SC | = +SC | = +SC | -ngi- | engim(u)- | engi- |
| 1p pl | — | si- | s- | s- | = +SC | = +SC | = +SC | -si- | esiba- | esi- |
| 2p sg | — | u- | w- | w- | = +SC | = +SC | = +SC | -ku- | om(u)- | o- |
| 2p pl | — | ni- | n- | n- | = +SC | = +SC | = +SC | -ni- | eniba- | eni- |

Note that +SC depends on whether the initial letter of what follows is a consonant (+C), an 'a' or 'e' (+a/e), or an 'o' (+o).

Meanings of the abbreviations in Tables 4 and 5

| Abbreviation | Meaning | Abbreviation | Meaning |
|--------------|-----------------------|--------------|-------------------|
| p | person | SC | subject concord |
| sg | singular | OC | object concord |
| pl | plural | AP | adjective prefix |
| + | positive | RC | relative concord |
| - | negative | C | consonant |
| ∅ | no concord | V | vowel |
| ~ | situative; continuous | N | nasal |
| § | subjunctive | VERB | verb root or stem |

Table 5: Verb formulas for the main Zulu tenses and moods

(Note that an OC, when present, is always found immediately before the verb root or stem.)

| Tenses and Moods | + | - | Formula | Examples | |
|-------------------|--------------|-----------------|---|---|---|
| Imperative | + | sg | VERB_a | <i>bhala!</i> 'write!', <i>hamba!</i> 'go!' | |
| | | pl | VERB_a_ni | <i>bhalani!</i> 'write, all of you!', <i>hambani!</i> 'go, all of you!' | |
| | VERB = VC... | { | sg | y_VERB_a | <i>yakha!</i> 'construct!', <i>yenza!</i> 'do!' |
| | | | pl | y_VERB_a_ni | <i>yakhani!</i> 'construct, all of you!', <i>yenzeni!</i> 'do, all of you!' |
| | VERB = C(C) | { | sg | yi_VERB_a | <i>yima!</i> 'stand!', <i>yidla!</i> 'eat!' |
| | | | pl | yi_VERB_a_ni | <i>yimani!</i> 'stand, all of you!', <i>yidlani!</i> 'eat, all of you!' |
| | - | sg | musa uku_VERB_a | <i>musa ukudlala!</i> 'don't play!', <i>musa ukwesaba!</i> 'don't be afraid!' | |
| | | pl | musani uku_VERB_a | <i>musani ukuhleka!</i> 'don't laugh, all of you!', <i>musani ukwethuswa!</i> 'don't be surprised, all of you!' | |
| Present | + | long | +SC_ya_VERB_a | <i>uZodwa uyakhala</i> 'Zodwa is crying', or 'Zodwa cries' | |
| | | | +SC_ya_OC_VERB_a | <i>ziyababulala</i> 'they are destroying <u>them</u> ' | |
| | | short | +SC_VERB_a | <i>igazi ligijima kalula</i> 'the blood is running easily' [+object] | |
| | - | | +SC_OC_VERB_a | <i>uma intombi ingithanda</i> 'if a girl loves <u>me</u> ' | |
| | | | a_-SC_VERB_i | <i>intsha yale mihla ayithandi ...</i> 'the youth of today do not like ...' | |
| | | a_-SC_OC_VERB_i | <i>umfundisi akabaphi amaswidi</i> 'the pastor doesn't give <u>them</u> sweets' | | |

| | | | | |
|------------------------------------|-------|------------------------|--|--|
| Perfect / Near past | + | long | +SC_VERB_ile | <i>lowo mfana ubalekile</i> 'that boy ran away' |
| | | | +SC_OC_VERB_ile | <i>ngik<u>u</u>tshelile</i> 'I've told <u>you</u> ' |
| | short | | +SC_VERB_e | <i>sizwe izindaba</i> 'we heard the news' |
| | | | [+object] +SC_OC_VERB_e [+object] | <i>uthisha ub<u>av</u>umele amaswidi</i> 'the teacher allowed <u>them</u> sweets' |
| - | | a_-SC_VERB_anga | <i>angilalanga</i> 'I could not sleep', or 'I didn't sleep' | |
| | | a_- SC_OC_VERB_anga | <i>wena awu<u>yiz</u>wanga le ndaba?</i> 'haven't you heard <u>this story?</u> ', <i>wena awu<u>yiz</u>wanga?</i> 'you have not heard <u>it?</u> ' | |
| Stative | + | | +SC_VERB_ile | <i>kubalulekile</i> 'it is important' |
| | - | | a_-SC_VERB_ile | <i>angilambile</i> 'I am not hungry' |
| Remote past | + | | +SC_a_VERB_a | <i>cishe ngafa!</i> 'I nearly died!' |
| | | | +SC_a_OC_VERB_a | <i>wan<u>g</u>itshela</i> 'you told <u>me</u> ' |
| | - | | a_-SC_a_VERB_a | <i>akahlala</i> 'he didn't sit' |
| Near future | + | | +SC_zo_VERB_a | <i>bazoqhubeka</i> 'they will continue' |
| | | | +SC_zo_OC_VERB_a | <i>ngizom<u>t</u>shela ubaba</i> 'I will tell <u>father</u> ', <i>ngizom<u>t</u>shela</i> 'I will tell <u>him</u> ' |
| | - | | a_-SC_zu_VERB_a | <i>abazali abazukhokha</i> 'the parents will not pay' |
| | | | a_- SC_zu_OC_VERB_a | <i>angizum<u>b</u>uza</i> 'I will not ask <u>her</u> ' |
| Remote future | + | | +SC_yo_VERB_a | <i>siyohlala</i> 'we shall stay' |
| | | | +SC_yo_OC_VERB_a | <i>umkako uyom<u>t</u>handa</i> 'you will love <u>your wife</u> ', <i>uyom<u>t</u>handa</i> 'you will love <u>her</u> ' |
| | - | | a_-SC_yu_VERB_a | <i>akuyubuya muntu</i> 'no one will return' |
| Present Situative / Participial | + | | ~SC_VERB_a | <i>ngababona ababili beguqa</i> 'I saw the two kneeling down' |
| | - | | ~SC_nga_VERB_i | <i>uma bengaphumeleli empilweni</i> 'if they are not succeeding in life' |
| Subjunctive | + | | §SC_VERB_e | <i>athathe amanzi aphuze</i> 'then she takes water and drinks' |
| | - | | §SC_nga_VERB_i | <i>intsha ingagcini</i> 'the youth must not stop' |
| Past continuous | + | | be_-SC_VERB_a | <i>bengihamba</i> 'I was walking' |
| Future continuous | + | | +SC_zo_be | <i>uzobe egijima</i> 'he will be running', <i>amakwaya</i> |
| | | | ~SC_VERB_a | <i>azobe ealana imilala</i> 'the choirs will be competing' |

Table 6: A selection of other verbal formatives, to be cut off before looking up Zulu verbs

| Formative | Meaning | Translation | Place in formula |
|--------------|--------------------|------------------------|------------------|
| -sa- | + progressive | 'still' | after SC |
| .. | - progressive | 'no longer' | .. |
| -nga- | + potential | 'can', 'may', 'could' | after SC |
| -nge- | - potential | .. | .. |
| -bo- | request | 'must' | after SC |
| -ka- | exclusive negative | 'not yet' | after SC |
| ka-, ma-, a- | hortatives | 'please', 'let' | before SC |
| -yo | relative | 'who', 'which', 'that' | after VERB |
| -ni | question | 'what?' | after VERB |

Comments 19: The verb being the hardest word class to parse in Zulu, Table 5, together with the supporting Tables 4 and 6, is truly the most important summary in the mini-grammar *as a grammar*, which should enable the dictionary users to decode (as well as encode) the language. These tables are undoubtedly overwhelming, a direct result of the highly complex nature of Zulu. No existing grammars of Zulu were consulted in building these tables. Rather, the Zulu corpus was queried and all frequent (and only the frequent) structures seen were described. The use of underscores to indicate morpheme boundaries in the verb formulas, as well as the decision to use a number of different labels for the subject concords (+SC +C, +SC +a/e, +SC +o, -SC, ~SC, §SC) all developed naturally during the effort to describe the corpus patterns seen. Rather than underscores, existing grammatical descriptions in Bantu at large also use dashes, hyphens and dots. Interestingly, for Zulu, the use of a well-defined set of subject concords was utilized to great effect in a series of studies on the Zulu verb published in the 1960s (Beuchat 1963, 1964, 1964a, 1966). In hindsight, the linguists who worked on Zulu half a century ago could very well have come up with overview tables such as Table 5. What they didn't have was evidence on frequency of occurrence. Also observe that all examples used in Table 5 have been selected from the OZSD, so are *real* examples. Parts in bold in the formulas and examples always refer to subjects, underlined parts to objects. While considerable efforts were put into the creation of this section of the mini-grammar, it should be remembered that this is *not* an exhaustive treatment of all aspects of the verb in Zulu, and as such far from complete. In places it is also decidedly too approximate. As a tool to support successful dictionary use, however, thus viewing the mini-grammar *as a guide to use the dictionary*, the level of detail is more or less the furthest one can go for a school dictionary.

The actual use of the material presented in the tables is illustrated next in the mini-grammar, followed by brief sections on auxiliary and copulative verbs.

For example, in the sentence *Izidakamizwa ziyababulala abadlali abancane* 'Drugs are destroying the young players', you should analyze the verb according to the formula +SC_ya_OC_VERB_a.

| | | | | |
|-------------------------|--------------------------------|----------------------------|------------------------|------------------------|
| word in sentence: | <i>izidakamizwa</i> | <i>ziya<u>babulala</u></i> | <i><u>abadlali</u></i> | <i><u>abancane</u></i> |
| entry in dictionary: | izidakamizwa > isidakamizwa | -bulala | abadlali > umdlali | abancane |
| word class + class no.: | pl. noun 7/8 | +SC8_ya_OC2_VERB_a | pl. noun 1/2 | adjective 2 |
| translation: | drugs | they are destroying them | players | young |

The *-be* in the formulas for the past and future continuous, seen in the last two blocks of Table 5, is actually the auxiliary verb stem *-be*. Please consult your Zulu textbook for the correct use of this auxiliary verb stem, as well as for the other frequently used auxiliary verb stem *-se*. Several usage examples may also be found in your dictionary, under the entries *-be* and *-se*. In addition to *-be* and *-se*, which are complex to use, there are also 50 other auxiliary verbs in your dictionary. Auxiliary verbs are typically used together with other verbs. For examples, see under *-azi²*, *-buye* or *-ngahle* in your dictionary.

Comments 20: In (9) and (10) below, the complete articles for the auxiliary verbs *-be* and *-azi²* as found in the OZSD are shown:

- (9) **-be** *** *auxiliary verb* 1 ■ was/were (in a state); used to; had; might/may/could/would have ♦ **Wabe** enezizathu zakhe ezamenza wesaba ukubuya emini. *He had his reasons to fear returning in the afternoon. (Literally: He was in a state of being with reasons ...)* ♦ **Babe** bephuma naye ngomnyango ongemuva. *They came out with him through the back door. (Literally: They were in a state of coming out ...)* 2 ■ will/shall (be in a state); will/shall have ♦ **Umqhudlwano uzokube** use-Stellenbosch. *The competition will be in Stellenbosch.* ♦ **Kuzobe** sekungokwesibili evela kule nkantolo ngala macala. *It will be the second time that he appears in this court in connection with these charges.*

This auxiliary verb actually indicates a continuous action or process, either in the remote past (sense 1) or in the future (sense 2). The meaning of this verb is not always reflected in English translations. Note that “-be” can also be the past tense of the copulative verb “-ba”.

- (10) **-azi²** *** *auxiliary verb* C -AZI¹ ■ be able to; know how; can ♦ **Abafundi bangakwazi** ukuzicabangela okuqhubekayo. *The students should be able to think for themselves progressively.*

The copulative verb stem *-ba* (with its past tense *-be*) is also complex to use. For examples, see your dictionary under *-ba*, but only your Zulu textbook will teach you the correct use.

Comments 21: There is just one copulative verb stem in Zulu, namely *-ba*, which is why it is this verb that ends the section on verbs. In (11) below, the OZSD article for *-ba* is shown:

- (11) **-ba** *** *copulative verb* ■ be; become ♦ **Wahleka kwaze kwaba** buhlungu izimbambo. *He laughed until his ribs became sore.* ♦ **Mina bengithi kungaba** ngcono lokho sikwenze kusasa. *I thought that it would be better if we did that tomorrow.*

3.6 Adjectives

In Zulu, there are only a few adjective stems (such as *-nye*; *-khulu*; *-bili*; etc.). In your dictionary, 17 are covered, and for each of those adjective stems all the frequent full forms have been listed. ‘Full form’ here means the **adjective stem** together with the **adjective prefix** (AP in Table 4). As such, adjectives like *omunye*, *ezinye*; *omkhulu*, *ezinkulu*; *ababili*, *ezimbili*; etc. can all be looked up directly, thus under the first letter of the adjective prefix.

Comments 22: For an exhaustive treatment of the lemmatization of adjectives in the OZSD, see De Schryver (2008a). In (12) and (13) below, two examples are given:

- (12) **omkhulu**¹ *** *adjective cl. 1, cl. 3* ■ big; large; great ♦ Ngithe ngilele ngaphupha ngibona umuntu **omkhulu** evele phakathi kwamafu. *Every time I dreamt, I saw a large person appearing between the clouds.*
◇ **uthisha omkhulu** ■ principal; headmaster ♦ UThisha **omkhulu** wafunda eNcwadini eNgcwele. *The principal read from the Holy Book.*
▶ **mkhulu** ■ master; head (*used to address someone respectfully*) ♦ Ngiyabonga, **mkhulu**. *Thank you, master.*
▶ **omkhulukazi** ■ very big; very large; very great; huge ♦ Inkosi yakwaButhelezi ibize khona umhlangano **omkhulukazi**, lapho wonke umuntu ongumZulu ebelindeleke khona. *The chief of the Buthelezi clan called a great meeting at which every Zulu person was expected.*

(For more on the suffix *-kazi*, as in *omkhulukazi*, see Gauton et al. 2004.)

- (13) **ezinkulu** ** *adjective cl. 8, cl. 10* ■ big; large; great ♦ Bayothenga ezitolo **ezinkulu** nasemakethe eku-Warwick Avenue eThekwini. *They are going to shop in the large stores and the market in Warwick Avenue in Durban.* ♦ “Awu! Kanti yinina zinsizwa **ezinkulu!**” *“My word! Is it you, big guys?”*

3.7 Relatives

Compared to adjective stems, there are more relative stems in Zulu (e.g. *-mhlophe*; *-mnandi*; *-thile*). In your dictionary, 26 are covered, and for each of those relative stems all the frequent full forms have been listed. ‘Full form’ here means the **relative stem** together with the **relative concord** (RC in Table 4). As such, relatives like *emhlophe*, *obumhlophe*; *emnandi*, *obumnandi*; *ethile*, *obuthile* can all be looked up directly (under the first letter of the relative concord).

Comments 23: By and large, the lemmatization of relatives in the OZSD runs parallel to the way in which adjectives have been lemmatized. This is also implicit in the heading of Table 4, where both are contrasted to the way in which verbs have been lemmatized. In (14) and (15) below, two examples are given:

- (14) **emnandi** *relative cl. 4, cl. 9* ■ nice; pleasant ♦ Maningi amaphethini emisindo **emnandi** angadaleka ngokuphindaphinda. *There are many pleasant sound patterns that can be created by repetition.* ♦ Wonke umuntu ufuna impilo **emnandi**. *Every person wants a nice life.*
- (15) **obumnandi** *relative cl. 14* ■ nice; pleasant ♦ Ngizolala ubuthongo **obumnandi** namhlanje. *I will have a nice sleep today.*

3.8 Adverbs

Zulu **adverbs** (for example *impela*, *kangaka*, or *nakanjani*) can be looked up directly.

Comments 24: Although the lemmatization of what Taljaard and Bosch (1993²: 41) call 'true adverbs' should be plainly straightforward in any Zulu dictionary, some lexicographers manage to omit the adverb category altogether from their Zulu dictionaries. For example, the adverb *impela* 'really; actually; truly; indeed' — the 124th most frequent lemma in the OZSD — has not been lemmatized in Mbatha's (2006) monolingual dictionary for Zulu. Instead, the infrequent noun *impéla* 'the real one' has been lemmatized, as Mbatha, following Nkabinde (1975), does not 'recognize' any word classes besides nouns, verbs, ideophones and interjections! Although not accepted by anyone else working on the Bantu languages, it is unfortunate that the Zulu National Lexicography Unit compiled their dictionary within this framework.¹² The OZSD's dictionary entry for *impela* is shown in (16) below:

- (16) **impela** *** *adverb* ■ really; actually; truly; indeed ♦ U-Maurice wenze isimemezelo ukuthi uyamthanda **impela** u-Minah.
Maurice announced that he truly loves Minah.

Some adverbs have been subdivided into **locative adverbs** (for example *kuwe*, *ngaphambi*, *phansi*), **adverbs of manner** (for example *kabi*, *ngokusemthethweni*, *njengenhlayenza*), **temporal adverbs** (for example *kudala*, *manje*, *ngomhla*), **instrumental adverbs** (for example *ngabo*¹, *ngathi*, *ngazo*¹), **associative adverbs** (for example *nabo*¹, *nathi*, *nazo*), **comparative adverbs** (for example *njengabo*, *nje-ngathi*, *njengazo*), and **interrogative adverbs** (for example *kangakanani*, *kanjani*, *nini*).

Comments 25: In addition to 'true adverbs', adverbs may also be derived from words in other word classes. Whenever the words from which such adverbs are derived have also been lemmatized as words in the OZSD, a cross-reference is included. One example each of the adverb subtypes included in the OZSD is shown in (17) to (23) below:

- (17) **kuwe** *** *locative adverb 2p sg* ⚡ WENA
■ at/to/from/in/on/... you ♦ Luchazani
usuku lwakusasa **kuwe**? *What does tomorrow mean to you?*
- (18) **ngokusemthethweni** *adverb of manner*
⚡ EMTHETHWENI ■ officially ♦ Bonke
bafungiswe **ngokusemthethweni** yiJaji,
uMnu. Vuka Tshabalala. *All were sworn in
officially by the Judge, Mr Vuka Tshabalala.*
- (19) **kudala** ** *temporal adverb* ■ long ago; in
the past; in the old days ♦ Ngakutshela
kudala ukuthi angifuni ukushada. *I told
you a long time ago that I don't want to get
married.*

- (20) **ngathi** *instrumental adverb* 1p pl **C** THINA
■ about us • Ngesaba ukubatshelela **ngathi**
sebezoze bazibonele. *I am afraid to tell*
them about us; they will eventually find out
for themselves. ■ with us • Okwamanje
umqeqeshi usebenza kakhulu **ngathi**
ukuze sikwazi ukushaya amagoli amaningi.
At the moment the trainer is working very
hard with us, so that we can score many
goals.
- (21) **nabo**¹ *** *associative adverb* **C** BONA
1 ■ (cl. 2) with (them); to (them) • Hamba
nabo! *Go with them!* ■ also (they); (they)
too; even (they) • Bavuma **nabo** ngaleyo
ndledlana. *They also agreed in that small*
way. **2** ■ (cl. 14) also (it); (it) too; even (it)
• Njengazo zonke izinhlobo zesichasiso
ubumnini **nabo** bunezakhi ezimbili. *Just*
like all the other kinds of qualificative, the
possessive also has two formatives.
- (22) **njengazo** *comparative adverb* cl. 8, cl. 10
C ZONA ■ (just) like (them) • Phela
ngiyazazi mina ukuthi sengimdala
njengazo nje izalukazi Baba. *Of course, I*
know that I am old now, just like the old
ladies, Father. • Angilitholanga ithuba
lokufunda **njengazo** zonke ezinye
izingane. *I didn't get the opportunity to*
study like all the other children.
- (23) **nini** *** *interrogative adverb* ■ when? • Ufike
nini lapha? *When did you arrive here?*
◇ -**noma nini** ■ whenever; anytime; any day
• Yinto ongangena esitolo uyithenge noma
nini. *It is something that you can go into a*
store for any time and buy.

As a result of the word-based approach, numerous adverbs actually found their way into a dictionary for the very first time. From the above, all of the following have for example not been lemmatized in Doke and Vilakazi (1953²): *kuwe* 'at/to/from/in/on/... you', *ngokusemthethweni* 'officially', *njengenhlayenza* 'like every other day; as usual', *ngomhla* 'on (the day)', *ngabo* 'about (them, whom), etc.', *ngathi* 'about us, etc.', *ngazo* 'about (them, which), etc.', ..., *nabo* 'with (them), etc.', *nathi* 'with us', *nazo* 'with (them), etc.', ...

3.9 Pronouns

All Zulu pronouns can be looked up directly in your dictionary. They have been grouped together in seven different categories. **Possessive pronouns** are formed by combining a **possessive concord** (which refers to what is possessed) with a **possessive stem** (which refers to the possessor). In Table 7, you would therefore combine a PC with a Pstem to obtain forms like *akhe*, *kwaso*, or *lwami*. All these forms can be looked up directly, but you need to understand the numbers in the notation. For example: 6+1 at *akhe* means that someone in class 1 'possesses' something in class 6; 15+7 and 16+7 at *kwaso* means that a noun in class 7 'possesses' a noun in class 15 or class 16; 11+1p sg at *lwami* means that I

(the first person singular) 'possess' a noun from class 11. Please look at these examples in your dictionary so you become familiar with this system.

Comments 26: The lemmatization of possessive pronouns in the OZSD has been treated in De Schryver and Wilkes (2008). In (24) below, one example is given:

- (24) **kwaso** * possessive pronoun 15+7 ■ its;
her/his ♦ Lesi saga sisazi
ngokusetshenziswa **kwaso**. We know this
proverb by its usage. 16+7 ■ of it; of
her/him ♦ Wamisa phezu **kwaso** inqwaba
enkulu yamatshe ekhona nanamuhla. On
top of it he erected a great heap of stones
that are there to this day.

(I.e., the *ukusetshenziswa* 'usage' (cl. 15) of the *isaga* 'proverb' (cl. 7), and *phezu* 'on top' (cl. 16) of it, with 'it' something in cl. 7, to be inferred from the wider context.)

Table 7: Pronouns in Zulu (formatives and full words)

| Class no. | Class prefix | PC | Pstem | DEM I | DEM II | DEM III | Abs Pron | Incl Q Pron | Excl Q Pron | Incl Num Pron "2" | Incl Num Pron "3" |
|-----------|--------------|-------|--------|-------|--------|---------|----------|-------------|-------------|-------------------|-------------------|
| 1 | um(u)- | wa- | -khe | lo | lowo | lowaya | ye(na) | wonke | yedwa | — | — |
| 1a | u- | ka- | | | | | | | | | |
| 2 | aba- | ba- | -bo | laba | labo | labaya | bo(na) | bonke | bodwa | bobabili | bobathathu |
| 2a | o- | | | | | | | | | | |
| 3 | um(u)- | wa- | -wo | lo | lowo | lowaya | wo(na) | wonke | wodwa | — | — |
| 4 | imi- | ya- | -yo | le | leyo | leya | yo(na) | yonke | yodwa | yomibili | yomithathu |
| 5 | i(li)- | la- | -lo | leli | lelo | leliya | lo(na) | lonke | lodwa | — | — |
| 6 | ama- | a- | -wo | la | lawo | lawaya | wo(na) | (w)onke | (w)odwa | (w)omabili | (w)omathathu |
| 7 | is(i)- | sa- | -so | lesi | leso | lesiya | so(na) | sonke | sodwa | — | — |
| 8 | iz(i)- | za- | -zo | lezi | lezo | leziya | zo(na) | zonke | zodwa | zo(zi)mbili | zo(zi)ntathu |
| 9 | i(N)- | ya- | -yo | le | leyo | leya | yo(na) | yonke | yodwa | — | — |
| 10 | izi(N)- | za- | -zo | lezi | lezo | leziya | zo(na) | zonke | zodwa | zo(zi)mbili | zo(zi)ntathu |
| 11 | u(lu)- | lwa- | -lo | lobu | lolo | lobuya | lo(na) | lonke | lodwa | — | — |
| 14 | ubu- | ba- | [-bo] | lobu | lobo | lobuya | bo(na) | bonke | bodwa | bobubili | [bobuthathu] |
| 15 | uku- | kwa- | [-khu] | lokhu | lokho | lokhuya | kho(na) | konke | kodwa | kokubili | kokuthathu |
| 16 | pha- | kwa- | [-khu] | lapha | lapho | laphaya | — | — | — | — | — |
| 17 | ku- | — | — | lokhu | lokho | lokhuya | kho(na) | konke | kodwa | kokubili | kokuthathu |
| 1p sg | — | (wa-) | -mi | — | — | — | mi(na) | wonke | ngedwa | — | — |
| 1p pl | — | (ba-) | -ithu | — | — | — | thi(na) | sonke | sodwa | sobabili | sobathathu |
| 2p sg | — | (wa-) | -kho | — | — | — | we(na) | wonke | wedwa | — | — |
| 2p pl | — | (ba-) | -inu | — | — | — | ni(na) | nonke | nodwa | nobabili | nobathathu |

Meanings of the abbreviations and fonts in Table 7

| Abbreviation | Meaning | Abbreviation | Meaning |
|------------------|---|--------------|--------------|
| PC | possessive concord | Abs | absolute |
| Pstem | possessive stem | Pron | pronoun |
| DEM | demonstrative | Q | quantitative |
| I, II, III | position I, II, or III | Incl | inclusive |
| <i>Italics</i> | not in dictionary; frequency too low | Excl | exclusive |
| <i>[Italics]</i> | not in dictionary; not attested in corpus | Num | numeral |

From the possessive pronouns one can also 'derive' **relativised possessive pronouns**. All the frequent ones have been listed directly into your dictionary; see for example *awakhe* (< *akhe*), *eyabo* (< *yabo*), or *okwethu* (< *kwethu*). Also in this category are words like *eze mfundo*, *ezemidlalo*, or *eze mpilo*.

Comments 27: In (25) and (26) below, two examples of relativised possessive pronouns are given:

- (25) **awakhe** *relativised possessive pronoun*
6+1 **AKHE** ■ her/his (own); hers/his
♦ Kanti **awakhe** amabutho alwa kanjani,
Nxumalo? *In fact, how did his own soldiers*
fight, Nxumalo?
- (26) **eze mfundo** * *relativised possessive pronoun*
IZINDABA + **IMFUNDO** ■ education;
educational affairs ♦ UNkk. Thembisile
Dlamini ovela eMnyangweni
Weze mfundo, uthe bayaluseka kakhulu
lolu hlelo. *Mrs Thembisile Dlamini, who is*
from the Department of Education, said they
support this programme wholeheartedly.

There are three types of **demonstrative pronouns**, positions I, II and III, depending on the 'distance' relative to the speaker ('here', 'there', and 'over there'). See Table 7 for all the main forms, and your dictionary for corresponding examples. When a demonstrative pronoun precedes a noun, the pre-prefix of the noun is dropped, and the demonstrative pronoun and the following noun are written as two words in the current orthography. The **adverbial formatives** *nga-*, *na-*, and *njenga-*, on the other hand, may be prefixed to the demonstrative pronouns, and thus written as one word. These formatives need to be cut off before looking up a demonstrative pronoun.

Comments 28: The OZSD being a corpus-driven dictionary, paradigms are not necessarily completed for the sake of completeness. In Table 7, Roman type is used to mark all forms frequent enough to be covered in the A-to-Z section of the OZSD. On the other hand, italics is used to mark those forms for which the frequency is too low to warrant inclusion in the A-to-Z section, while italics plus square brackets is used for forms not attested in the corpus at all, and thus most definitely absent from the A-to-Z section. Frequent variants for position I, on the other hand, have been lemmatized.

In (27) to (30) below, one set is given for positions I, II and III for the class 4 and 9 demonstratives, as well as the corresponding variant for position I:

- (27) **le**¹ *** *demonstrative pronoun pos. I* **1** ■ (cl. 4)
these (ones) ♦ **Le** mithetho izoletha ezayo
izinkinga. *These laws will bring their own*
problems. **2** ■ (cl. 9) this (one) ♦ Iyiphi
imoto edale **le** ngozi? *Which car caused*
this accident?
- (28) **leyo** *** *demonstrative pronoun pos. II* **LE**¹
1 ■ (cl. 4) those (ones) ♦ **Leyo** misebenzi
yethu isemqoka njengayo yona **leyo**
eminye. *Those jobs of ours are as important*
as those other ones. **2** ■ (cl. 9) that (one)
♦ Kungabe yini **leyo**? *What is that one?*

- (29) **leya** *demonstrative pronoun pos. III* **C** L.E¹
1 (cl. 4) those (ones) • Umuntu omunye ubengathi sithandaza sibusisa **leya** mikhuba ebigilwa yilaba bafana uyazi? *You know, another person might think we are praying while blessing the tricks played by those boys.* **2** (cl. 9) that (one) • **Leya** ndawo iyaziwa ngomlando wayo. *That place is known for its history.*

The person(s) or object(s) referred to is/are far away in time or distance.

- (30) **lena** *demonstrative pronoun pos. Ia* **C** L.E¹
1 (cl. 4) these (ones) • Imikhonto **lena** ibazwa ngendlela ethize ebalulekile kulesi sizwe. *These assegais are shaped in a very special way in this nation.* **2** (cl. 9) this (one) • Indatshana **lena** ingahlukaniswa ibe yizingxenye eziyisithupha. *This short story can be divided into six parts.*

This pronoun always follows the noun.

In Zulu, the **absolute pronouns** are used for emphasis or contrast. The full forms shown in Table 7 have been listed in your dictionary, together with important usage notes for each of them. Note that a large number of adverbs may be derived from absolute pronouns (for instance: *bona* > *kubo*, *kubona*, *nabo*, *ngabo*, *njengabo*).

Comments 29: In (31) below, one example of an absolute pronoun is given:

- (31) **sona** *absolute pronoun cl. 7* **it** (in particular); she/he (in particular) • **Sona** isizwe sakwaZulu siyobuswa ngubani? *By whom will the Zulu nation be governed?* • **Yiso** isipho sakho leso, ngiyakupha. *That is your gift; I am giving it to you.*

The meaning of this pronoun is not always directly visible in an English translation. When "**sona**" is preceded by formatives such as "yi-", "si-", "na-" or "ku-", the final emphasizing syllable "-na" is usually deleted. See the second example.

Finally, there are also three types of quantitative pronouns in Zulu, with which quantities are expressed. All the forms that have been listed in your dictionary are shown in the last four columns of Table 7. The **inclusive quantitative pronouns** (stem *-nke*) mean 'the whole' in the singular and 'all' in the plural. The **exclusive quantitative pronouns** (stem *-dwa*) mean 'alone' or 'only'. The **inclusive numeral pronouns** (only frequent stems *-bili* and *-thathu*) are used to refer to groups of items (here 'both' and 'all three' respectively).

Comments 30: For an exhaustive treatment of the lemmatization of the three types of quantitative pronouns in the OZSD, see De Schryver (2008b). In (32) to (34) below, an example each is given:

- (32) **bonke** *** *inclusive quantitative pronoun*
1 ■ (cl. 2) all • Bawethamela **bonke** abantu umngcwabo. *All the people attended the funeral.* 2 ■ (cl. 14) all; the whole • Zilinde **bonke** ubusuku. *They waited the whole night.*
- (33) **odwa** *exclusive quantitative pronoun cl. 6*
Compare wodwa ■ alone; on their/its own
• Wase ebona amaphoyisa asehlangana **odwa** endleleni. *And then he saw the police assembling on their own in the street.* ■ only
• Ukhuluma amanga **odwa** wena. *You are only telling lies.*
- (34) **sobabili** *inclusive numeral pronoun 1p pl*
■ both of us • Uthishanhloko ubefuna ukusibona **sobabili** emva kwesikole. *The principal wanted to see both of us after school.*
◇ [NEG+] **sobabili** ■ neither of us

3.10 Ideophones

Ideophones as a word class do not exist in English. In a language such as Zulu they are highly characteristic and their use signals good and beautiful Zulu. There are 100 ideophones in your dictionary. Examples are *nya*, *phecelezi*, and *qho*. Ideophones have been described as “marked words that vividly evoke sensations and perceptions”. They are not problematic to look up, as no other morphemes are attached to them. However, they are **very hard to translate**. For that reason, only general paraphrases are given, between brackets and always preceded by ‘of ...’. Often, several examples are shown, to illustrate the range of possibilities. Only when combined with verbs and nouns can rather fixed meanings be given. See for example the combinations under *ngci*, *phaqa*, or *qakala*.

Comments 31: For an exhaustive treatment of the lemmatization of ideophones in the OZSD, see De Schryver (2009). In (35) and (36) below, two examples are given:

- (35) **nya** ** *ideophone* ■ (of complete absence)
• Kwathula kwathi **nya**. *It was completely silent.* • Wayefuna ukukhohlwa **nya** yile ndaba kaNomusa. *He wanted to forget this Nomusa affair completely.* • Inyoka uyishaye wayibulala **nya**. *He hit the snake and killed it altogether.*
- (36) **ngci** *ideophone* 1 ■ (of tightness, of density)
• Ithi **ngci** intuthu. *The smoke is very dense.*
• Abantu bacinana kuthi **ngci**. Akube kusaphefumuleka nakahle. *There is a great crowd; it is no longer possible to breathe properly.* 2 ■ (of extreme action) • Avuke umsindo uloku uthe **ngci**. *Then he woke up while the noise continued to be overwhelming.*

- ◇ **-gcina ngci** ■ be the very last • U-Ayanda ingane yokugcina **ngci** ekhaya. *Ayanda is the last-born at home.* • Uzocina **ngci ngci!** *It will be your very last time!*
- ◇ **-vala ngci** ■ close tightly • Wavala ifasitela **ngci.** *She closed the window tightly.*

The description of ideophones as “marked words that vividly evoke sensations and perceptions” in the mini-grammar is taken from Dingemans (2009). According to Blench:

Ideophones are abundant in natural and heightened speech, notably in Africa, but absent from typical example sentences, hence their failure to be treated adequately in typical grammars and dictionaries. They can be difficult to elicit since their existence is unpredictable and speakers have no natural ‘hook’ to recall them. Their elusive nature, in grammatical terms, has made them poor relations to other word classes and they have been little treated by the schools of grammar dominated by syntax [...] Our understanding of the role they play in natural language (as opposed to elicited examples) is still very preliminary.

— Blench (2009: 1)

As I pointed out in my own study of ideophones, compiling entries for ideophones in the OZSD “took an average three times longer than the compilation of entries in any other word class” (De Schryver 2009: 38). While the lemmatization proper of ideophones does not pose any problems (stable orthographic forms can simply be used as lemmas), I can reconfirm, with Blench, that the analysis and synthesis of large amounts of *natural language data* was simply paramount in order to make any sense at all of ideophones. Without the corpus used, in other words, it would simply not have been possible to treat ideophones adequately.

3.11 Conjunctions

Conjunctions in Zulu, as in English, introduce or link sentences. Examples are *bese*, *ukuze*, and *nxá*. Most conjunctions can be looked up directly, without the need to cut off additional prefixes. Conjunctions are very frequent. The most frequent word in Zulu is a conjunction (*ukuthi*), as well as the third-most frequent (*uma*), and the sixth-most frequent (*ngoba*).

Comments 32: In (37) below, the article for the conjunction *ukuthi* is shown:

- (37) **ukuthi** *** conjunction ☞ -THI' ■ that
• Kwezwakala **ukuthi** useshonile. *We heard that he passed away.* • Ngawatshela **nokuthi** u-Matthew wayekhona naye. *I also told them that Matthew was there as well.*
• Bebejatshuliswa **wukuthi** yimi owenze lo msebenzi omkhulu. *They were happy that it was I who did this great job.*

Ukuthi is one of the conjunctions to which morphemes may be prefixed. As is generally the approach in the OZSD, the top orthographic forms are illustrated in the examples. To that end, the orthographic forms cum linked corpus frequencies shown in (38) were available to the compilers. Although the frequency of *ukuthi* itself (131 950) is many times higher than the frequency of all the other forms together (5 527 + 4 552 + ...), in order to illustrate the conjunctive potential for this lemma, the next few forms (here two) were also selected for illustrative purposes in (37).

- (38) ukuthi <131950>, nokuthi <5527>, wukuthi <4552>, yokuthi <1517>, ngukuthi <774>, wokuthi <738>, ukuth <465>, elokuthi <315>, kungukuthi <315>, yikuthi <210>, kuwukuthi <141>, kunokuthi <131>, kwazisukuthi <90>, nawukuthi <62>, kwakungukuthi <44>

3.12 Copulatives

Copulatives in Zulu are typically **derived from words in other word classes**. To derive them, the prefixes are changed, with the goal to add the meaning 'is/are ...'. Derived from the relative stem *-mhlophe*, for example, the frequent copulatives listed in your dictionary are *imhlophe*, *umhlophe*, and *zimhlophe*. From the relative stem *-mnandi*, the frequent copulatives *amnandi*, *innandi*, *kumnandi*, and *zinnandi* are derived and listed. Derived from the enumerative stem *-phi*, your dictionary lists the following frequent copulatives: *(y)imuphi*, *yibaphi*, *yimiphi*, *(y)iliphi*, *yimaphi*, *(y)isiphi*, *yiziphi*, *(y)iyiphi*, *yiluphi*, and *(y)ikuphi*. And derived from the interrogative stem *-phi*, your dictionary lists the following frequent copulatives: *uphi*, *baphi*, *i(ku)phi*, *liphi*, *aphi*, *siphi*, *ziphi*, *luphi*, and *(ku)kuphi*. In each case, you can look up all these forms directly, as full words in the dictionary, where you will find precise grammatical information, including a mention of the class numbers, as well as specific examples.

Comments 33: No other dictionaries for Zulu lemmatize copulatives. The decision to consistently lemmatize words wherever possible in the OZSD, however, naturally led to their inclusion. In (39) below, an example is shown:

- (39) **umhlophe** *copulative 1* ■(cl. 1) (she/he) is white • Mfowethu, uma ungibukisisa ungashe ukuthi ngimnyama, yena umlungu **umhlophe** ngempela? *My brother, if you look carefully at me, would you really say that I am black and that the white man is white?* **2** ■(cl. 3) (it) is white • Umyeko wezinwele **umhlophe** wu ngubuhlalu. *The hair of the diviner is sparkling white with beads.* **3** ■(2p sg) you are white • Umuhle kunezihlabathi zolwandle / **Umhlophe** kunobisi lwezimazi zakwethu [imigqa emi-2 yenkondlo] *You are prettier than the sands of the sea / You are whiter than the milk of our cows [two lines of poetry]*

3.13 Interjections

Interjections in Zulu, as in English, are words 'thrown in' to mark surprise, pleasure, amazement, disbelief, disappointment, (dis)approval, (dis)agreement, etc. Examples are *awu*, *hhayi*, and *wo*. All interjections can be looked up directly.

Comments 34: In (40) below, an example of an interjection is shown:

- (40) **awu** *** *interjection* C HAWU ■ my word!;
good heavens!; look!; hey!; wow!; oh!;
eish! (*marks a surprise or disbelief*)
♦ Waphinde washaya izandla wathi: “Awu!
Awu! Awu! Usuyasazi impela isiZulu.” *She
clapped her hands again and said: “Wow!
Wow! Wow! You know Zulu very well.”* ♦ Awu
bayahlupha impela! *Eish, they are really
bothersome!*

3.14 Other word classes

Smaller word classes include **enumeratives** (for example all frequent forms of *-ni*: *muni*, *mini*, *lini*, *mani*, *sini*, *zini*, *yini*, *luni*, *buni*, and *kuni*), **locative demonstrative copulatives** (for example for class 1, positions I, II and III: *nangu*, *nango*, and *nanguya*), **possessives** (see for example *alo*², *balo*¹, *kwalo*², etc.), **interrogatives** (see for example *na*, *ngabe*², or *yini*²), so-called **suffixes** (for example *jikelele*, *ndini*, or *yansondo*), **enclitics** (see for example *bo*, *-ke*¹, or *nje*¹), **hortative markers** (*ake* and *make*), and **abbreviations** (*a.m.* and *p.m.*).

Comments 35: In (41) to (48) below, an example for each of these word classes is shown:

- (41) **muni** *enumerative cl. 1, cl. 3* ■ what kind of?
♦ Muntu **muni** ongathandi ukusizwa? *What
kind of person doesn't like to be helped?*
♦ U-Vincent wenza msebenzi **muni**? *What
kind of work does Vincent do?*
- (42) **nangu** ** *locative demonstrative copulative
pos. I cl. 1* ■ here (she/he) is ♦ **Nangu**
umuntu eqeda imali yabantu. *Here is the
person who wasted the people's money.*
This word can also be used as a demonstrative,
with the meaning “this”, e.g. “**nangu**
umfowethu” = “this brother of mine/ours”.
- (43) **alo**² *possessive C LO 1 6+1* ■ of/for this
♦ Elinye lamathemba **alo** Marion Jones
owawina izindondo ezinhlanu e-Sydney,
e-Australia ... *One of the hopes of this Marion
Jones who won five medals in Sydney,
Australia ...* 2 6+3 ■ of/for this
♦ Amalungiselelo e-Nigeria **alo** mdlalo
aphazanyiswe wukulimala kukakaputeni
yayo. *Nigeria's preparations for this game
were interrupted by the injury of their
captain.*

(For the next three entries, no translation equivalents are available. Note how the *function* is described rather, starting with the formulaic “marks ...”.)

- (44) **na** *** *interrogative* ■ (*marks interrogative
sentences*) ♦ Usaphila **na**? *Are you fine?*
♦ Kwenzenjani **na** khona? *What happened
there?* ♦ Ungizwa kahle **na**? *Do you
understand me?*

- (45) **ndini** * suffix ■ (marks sarcasm or contempt)
♦ Ake uthule Ndondo **ndini**, akusiwe ophethe lo mhlango. *Please keep quiet, smart Ndondo, you are not in charge of this meeting.* ♦ Ziphukuphukundini, niyedelela kanti. *You stupid fools, you are rude after all.*

Although “**ndini**” is a noun suffix, it is generally written separately.

- (46) **bo** *** enclitic ■ (marks insistence or anger)
♦ Yicala lani-ke leli osulenzile? Khuluma **bo!** Ubulale umuntu yini? *What type of crime have you committed? Speak up! Did you kill the person or not?* ♦ Hhayi **bo!** No, dammit!
- (47) **ake** *** hortative marker ■ will you (please) ... ; let ... (please) (marks a polite request)
♦ Ake uchaze isifundo esitholakala kule ndaba. *Will you please explain the lesson found in this story?* ♦ Manje **ake** sibheke le misho sibone ukuthi izenzo zakhiwa kanjani. *Now let's look at these sentences and see how the verbs are structured.*
- (48) **a.m.** abbreviation ■ a.m.; in the morning (any time between midnight and midday)
♦ Inkonzo izoqala ngesikhathi sika-10:00 **a.m.** *The church service will begin at 10:00 a.m.*

Sehora le-10 ekuseni is the traditional way to express at ten in the morning.

4. Discussion

Despite the fact that this paper is now already twice as long as the average scientific paper, the analysis of the OZSD is not all-embracing. This could not have been otherwise, as about a thousand pages are needed to do so (cf. Section 2). It is hoped, however, that all the claims made in the OZSD's Introduction have now been sufficiently substantiated:

The Zulu mini-grammar teaches you where to find particular words in the Zulu to English side of this dictionary, and thus teaches you *Zulu-specific dictionary skills*. It is very important that you study this section, because the method used to list Zulu words in this dictionary is new and therefore unfamiliar, but certainly more user-friendly. The result is a new type of Zulu dictionary, for the following reasons. **Complete meaningful words** have been entered in this dictionary, rather than parts of words. The selection of Zulu headwords is thus unique to this dictionary. The **modern class numbers** are used for all headwords: for nouns, of course, but also for all other word classes (parts of speech) that need to be in harmony with the nouns they refer to. **Informative cross-references** not only link verb stems with verb roots, but also derived nouns with the verbs they are derived from. In addition to headwords selected for their high frequency, this dictionary also treats all **frequent combinations** and **frequent derivations**. Headwords, combinations and derivations are illustrated with **authen-**

tic Zulu examples, taken from a **large corpus** of sentences that have actually been written or spoken before. A corpus consists of hundreds and hundreds of texts, containing millions of words, that have been taken from both the general language and from school textbooks. All **core and current meanings** have been listed, based on such corpus evidence.

— De Schryver (2010: xi), emphasis in bold as in original

The actual reception of the OZSD will ultimately be the litmus test for the claim that this dictionary is more user-friendly as a school dictionary than any other existing dictionary for Zulu. I had noticed (see e.g. De Schryver 2008a: 64) that the stem approach to the lemmatization of Zulu failed this particular target user group, so I set out to develop a daring word approach. In the process I also introduced the various other novelties marked in bold in the quote from the Introduction above — all of them used for the first time in Zulu lexicography. It is important to recall, however, that this new approach was specifically designed for young learners. Although I am indeed convinced that it is ideal for them, I by no means want to claim that I have solved all look-up problems for a highly conjunctive Bantu language such as Zulu. By lemmatizing all word classes except one as words, I basically bring the entire look-up problem back to recognizing and dealing with the one remaining word class, viz. verbs. Reformulated, in decoding Zulu, the one remaining orthographic word in a sentence after the easier ones will have been looked up, will be the verb. One thus also knows it must be a verb, at which point Tables 4 to 6 can be unleashed. The verb is also the only word class in the OZSD that is lemmatized with a preceding dash, indicating that something was cut off before reaching the lemma in the dictionary.¹³

Compared to stem lemmatization, word lemmatization is undoubtedly more repetitive, even though that repetition is tailored to each sense of each lemma anew. The information that is packed in a stem dictionary has been unpacked in a word dictionary. Some level of generalization is therefore missed, though one could argue that providing that is the task of a grammar, not a dictionary. Up to a point, the mini-grammar restores this, and repacks. But the mini-grammar is not complete: a section on the morphophonological (sound) changes, for example, could have been added had there been space for it. What word lemmatization does do is to put the lexicon centre-stage, with the intricacies of each word dealt with in detail. The lexicon is seen as the pivot in mastering a language, not the grammar. Grammar can be built around the lexicon. This is quite a reversal of Bloomfield's (1933: 274) view of "[t]he lexicon [as] an appendix of the grammar, a list of basic irregularities". Being able to approach Zulu words rather than Zulu roots and stems is a direct result of the corpus revolution. So are most other microstructural innovations, chief among those the authentic examples to illustrate the synthesised analysis.

Abandoning generalizations, in combination with the extra information categories offered in the OZSD, also means that the focus shifts from quantity to quality. Rather than offering 30 000 lemmas on about 920 pages as in Doke and Vilakazi's *Zulu-English Dictionary* (1953²), or 13 600 lemmas on about 220

pages as in Dent and Nyembezi's *Scholar's Zulu Dictionary* (1995³), the OZSD offers 5 000 lemmas on about 270 pages.¹⁴ The OZSD does fill a gap in the market, however, both in terms of its user-friendly access structure, and in terms of its coverage. Microstructurally, grammar was in effect brought into each dictionary article, systematic exemplification is a first for Zulu, and of course the lexicon was brought in line with current usage: meaning shifts resulted in new meanings which are now recorded, and hundreds of 'new' words were added to the macrostructure.

This paper comes after the lemmatization of some Zulu word classes have already been treated in the literature, and it is hoped that it can serve as a launching pad for all the remaining ones. For those treated before (possessive pronouns, adjectives, quantitative pronouns, and ideophones) comparing the tiny summaries presented in the mini-grammar (respectively Sections 3.9, 3.6, 3.9, and 3.10 above) gives an idea of how a scientific description may be presented to the dictionary user. Conversely, the summaries for the other word classes give an idea of what the scientific descriptions will conclude. Concluding is one level, going through the details another, and performing the actual corpus analysis — in order to first synthesize the facts to then compile the dictionary articles themselves — is yet another, far more complex, level. In order for a thriving Bantu metalexigraphy to develop, more dictionaries will have to be compiled for these languages, so that more dictionary compilers will be able to share their experiences. As such, the present contribution is but one such attempt, hopefully one that will stimulate debate.

Endnotes

1. A workbook to accompany the Zulu dictionary is being planned as well.
2. While compiling the Northern Sotho dictionary took slightly over two and a half years, for the Zulu one nearly four years were required. From a lexicographic point of view, a conjunctively-written language such as Zulu is indeed much more complex to handle than a disjunctively-written one such as Northern Sotho.
3. With GET the general education and training band. Grades 10 to 12 (the last three years of secondary education) are known as FET, the further education and training band.
4. Together, the two A-to-Z sections and all extra-matter texts add up to 640 pages, or thus in the book-binding jargon, twenty 32-page signatures ($20 \times 32 = 640$). Although tweaking the different sections, as well as the contents of those sections, until such a round multiple is reached is an important aspect of finalising an actual dictionary for the trade, this aspect will not be covered here.
5. Sincere thanks are due to A. Wilkes and D. Gowlett for their critical evaluation of the mini-grammar, and to M. Hall for making sure the English of the mini-grammar is on the level of the intended target user group.
6. Verb lemmas, even as roots or stems, are actually also orthographic words, but only in their imperative forms. As such, one could say that the OZSD managed to lemmatize *all* Zulu words as words. However, given verbs rarely occur in their imperative forms, it is more cor-

rect to stick to the traditional terminology, and to admit that verbs are lemmatized as roots and stems.

7. For more information on the preparation of a tailored lemma-sign list for junior dictionary users, balancing a general and a customised sub-corpus, see De Schryver and Prinsloo (2003).
8. The contract to produce the OZSD was signed between Oxford University Press Southern Africa and TshwaneDJe HLT, makers of the dictionary production system TLex.
9. Ideally, one would also have known which Zulu textbooks the learners would be using, so as to avoid the possibility of a diverging terminology, but in the absence of that knowledge, the best that could be done was to define each term and to use it accordingly in the OZSD.
10. Considerable thought also went into the way to present the information shown in Table 1. One adjudicator in particular, suggested using dedicated columns for mono- vs. polysyllabic stems, additional columns for vowel- vs. consonant-initial stems, and then another column for the remainder of the notes now in the last column. I feel that a single notes column, with numbered notes referring to the earlier columns, is sufficiently clear.
11. In an electronic environment, where space is not an issue, lemmatizing all frequent verb forms becomes a possibility, either with the help of unlimited human resources or computational techniques, but at that point other problems surface, such as the abandonment of important generalizations (cf. De Schryver 2008: 269-270).
12. For a brief analysis of Mbatha's (2006) dictionary, and its implications, see De Schryver and Wilkes (2008: 829-830).
13. There are only seven exceptions in the OZSD: the conjunctions *-thi* and *-the* (unique in that these two only can take verbal prefixes), the relative stems *-mbumbulu*, *-thize*, and *-thizeni* (as these three are always used in compounds), the relative / copulative *-emqoka* (the only such combined word class), and the enclitic *-ke* (always attached to another word, with dash).
14. Extrapolated lemma-sign counts for Doke and Vilakazi as well as Dent and Nyembezi are taken from De Schryver (2009: 52).

References

- Bennett, P.R.** 1986. Grammar in the Lexicon, Two Bantu Cases. *Journal of African Languages and Linguistics* 8(1): 1-30.
- Benson, T.G.** 1964. A Century of Bantu Lexicography. *African Language Studies* 5: 64-91.
- Beuchat, P.-D.** 1963. A Restatement of the Zulu Verb Conjugation. Part 1. *African Studies* 22(4): 137-169.
- Beuchat, P.-D.** 1964. A Restatement of the Zulu Verb Conjugation. Part 2. *African Studies* 23(1): 35-49.
- Beuchat, P.-D.** 1964a. A Restatement of the Zulu Verb Conjugation. Part 3. *African Studies* 23(2): 67-87.
- Beuchat, P.-D.** 1966. The Perfect Tenses in Zulu. *African Studies* 25(2): 61-71.
- Blench, R.** 2009. *The Sensory World; Ideophones in Africa and Elsewhere*. Draft circulated for comment; 5 October 2009 [online]. <http://www.rogerblench.info/Language%20data/Africa/General/Blench%20African%20Ideophones.pdf>
- Bloomfield, L.** 1933. *Language*. New York: Henry Holt & Co.
- Dent, G.R. and C.L.S. Nyembezi.** 1995³ [1988², 1969]. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.

- De Schryver, G.-M.** 1999. *Bantu Lexicography and the Concept of Simultaneous Feedback, Some Preliminary Observations on the Introduction of a New Methodology for the Compilation of Dictionaries with Special Reference to a Bilingual Learner's Dictionary Cilubà-Dutch*. Unpublished MA Dissertation. Ghent: Ghent University.
- De Schryver, G.-M.** 2001. *Pukuntšuthaloši ya Sesotho sa Leboa 1.0 (PyaSsaL's First Parallel Dictionary)*. Pretoria: (SF)² Press.
- De Schryver, G.-M.** 2004. *Concepts and Tools for Lexicography in the Electronic Age — A Case Study of Dictionary Compilation in South Africa*. Unpublished PhD Thesis. Ghent: Ghent University.
- De Schryver, G.-M.** 2006. Compiling Modern Bilingual Dictionaries for Bantu Languages: Case Studies for Northern Sotho and Zulu. Corino, E., C. Marellò and C. Onesti (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6-9 settembre 2006 / Proceedings XII Euralex International Congress, Torino, Italia, September 6th-9th, 2006*: 515-525. Alessandria: Edizioni dell'Orso.
- De Schryver, G.-M.** 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšu ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.
- De Schryver, G.-M.** 2008. Why Does Africa Need Sinclair? *International Journal of Lexicography* 21(3): 267-291.
- De Schryver, G.-M.** 2008a. A New Way to Lemmatize Adjectives in a User-friendly Zulu-English Dictionary. *Lexikos* 18: 63-91.
- De Schryver, G.-M.** 2008b. The Lexicographic Treatment of Quantitative Pronouns in Zulu. *Lexikos* 18: 92-105.
- De Schryver, G.-M.** 2009. The Lexicographic Treatment of Ideophones in Zulu. *Lexikos* 19: 34-54.
- De Schryver, G.-M.** 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: IsiZulu NesiNgisi, Esishicilelwe abakwa-Oxford*. Cape Town: Oxford University Press Southern Africa.
- De Schryver, G.-M. and R. Gauton.** 2002. The Zulu Locative Prefix ku- Revisited: A Corpus-based Approach. *Southern African Linguistics and Applied Language Studies* 20(4): 201-220.
- De Schryver, G.-M. and N.S. Kabuta.** 1997. *Lexicon Cilubà-Nederlands, Een circa 2500-lemma's-tellend strikt alfabetisch geordend vertalend aanleerderslexicon met decodeer-functie ten behoeve van studenten Afrikaanse Talen & Culturen aan de Universiteit Gent*. Recall Linguistics Series 1. Ghent: Recall.
- De Schryver, G.-M. and D.J. Prinsloo.** 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — with Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*: 216-242, 372. Dar es Salaam: TUKI, Chuo Kikuu cha Dar es Salaam.
- De Schryver, G.-M. and D.J. Prinsloo.** 2003. Compiling a Lemma-sign List for a Specific Target User Group: The Junior Dictionary as a Case in Point. *Dictionaries: Journal of The Dictionary Society of North America* 24: 28-58.
- De Schryver, G.-M. and A. Wilkes.** 2008. User-friendly Dictionaries for Zulu: An Exercise in Complexicography. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15-19 July 2008*: 827-836. Sèrie Activitats 20. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

- Dingemans, M.** 2009. *How to Do Things with Ideophones: Observations on the Use of Vivid Sensory Language in Siwu*. Presentation at the SOAS Research Seminar, June 3. London: The School of Oriental and African Studies.
- Doke, C.M. and B.W. Vilakazi.** 1953² [1948]. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Gauton, R., G.-M. de Schryver and L. Mohlala.** 2004. A Corpus-based Investigation of the Zulu Nominal Suffix -kazi: A Preliminary Study. Akinlabi, A. and O. Adesola (Eds.). 2004. *Proceedings of the 4th World Congress of African Linguistics, New Brunswick 2003*: 373-380. Cologne: Rüdiger Köppe Verlag.
- Mbatha, M.O.** 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.
- Nabirye, M.** 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished MA Dissertation. Kampala: Makerere University.
- Nabirye, M.** 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwegwa Olusoga* [A Dictionary of Lusoga. For Speakers of Lusoga, and for Those Who Would Like to Learn Lusoga]. Kampala: Menha Publishers.
- Nabirye, M.** 2009a. Compiling the First Monolingual Lusoga Dictionary. *Lexikos* 19: 177-196.
- Nabirye, M.** 2009b. Dictionary Compilation for Mother-tongue Speakers of Bantu Languages. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary & Lexicology*: 597-607. Beijing: Beijing Normal University.
- Nkabinde, A.C.** 1975. *A Revision of the Word Categories in Zulu*. Unpublished PhD Thesis. Pretoria: UNISA.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Special Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Prinsloo, D.J.** 2010. Review: *The Oxford Bilingual School Dictionary: Zulu and English*. *Lexikos* 20: 760-766.
- Prinsloo, D.J. and G.-M. de Schryver.** 2002. Towards an 11 x 11 Array for the Degree of Conjunctivism / Disjunctivism of the South African Languages. *Nordic Journal of African Studies* 11(2): 249-265.
- Schadeberg, T.C.** 1992³ [1984²]. *A Sketch of Swahili Morphology*. Grammatical Analyses of African Languages 2. Cologne: Rüdiger Köppe Verlag.
- Snoxall, R.A.** 1965. Some Problems and Principles of Lexicography in Luganda. *African Language Studies* 6: 27-31.
- Taljaard, P.C. and S.E. Bosch.** 1993² [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.
- Taljaard, E., G.-M. de Schryver, M. Mogodi and D. Paizee.** 2008. *Oxford Bilingual School Dictionary Workbook: Northern Sotho and English / Puku ya Mešomo: Pukuntš'u ya Polelopedi ya Sekolo: Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford*. Cape Town: Oxford University Press Southern Africa.
- TLex (formerly TshwaneLex).** 2010. TshwaneLex Dictionary Production System [online]. <http://tshwanedje.com/tshwanelex/>
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.

L'ingénierie lexicale ou la description d'un objet entre l'invention et la découverte*

Jean-Nicolas De Surmont, *Université de Leeds, Leeds, Royaume-Uni*
(jdesurmont@yahoo.fr)

Résumé: Résoudre des problèmes théoriques constitue également résoudre des problèmes lexicaux. C'est ce que nous avons baptisé *l'ingénierie lexicale*, processus d'évaluation et de description d'une activité ou d'une discipline et de création de termes ou de lexies pouvant satisfaire les besoins théoriques du chercheur. La nécessité de recourir à l'ingénierie lexicale peut apparaître dans différents contextes scientifiques notamment les études transdisciplinaires, les études diachroniques, l'harmonisation interlinguistique, etc. Nous présentons donc différents cas de figures notamment dans le cadre de nos travaux sur la poésie vocale.

Mots-clés: INGÉNIERIE LEXICALE, RÉOLUTION DE PROBLÈME (LEXIQUE), LEXICOLOGIE, TERMINOLOGIE, ÉPISTÉMOLOGIE

Abstract: **Lexical Engineering or the Description of an Object between Invention and Discovery.** Solving theoretical problems also involves solving lexical problems. This we have named *lexical engineering*, a process of evaluation and description of an activity or a discipline and the creation of terms or lexical items, which have to satisfy the theoretical needs of the researcher. The necessity to resort to lexical engineering becomes apparent in different scientific contexts, notably transdisciplinary studies, diachronic studies, interlinguistic harmonization, etc. We therefore present different kinds of figurative senses notably as part of our work on vocal poetry.

Keywords: LEXICAL ENGINEERING, SOLUTION OF (LEXICAL) PROBLEMS, LEXICOLOGY, TERMINOLOGY, EPISTEMOLOGY

L'ingénierie lexicale: définition

Loin de nous l'idée de défendre une quelconque hypothèse nominaliste ou réaliste du langage semblable aux célèbres thèses soutenues par Cratyle et Hermogène dans Platon si ce n'est que de montrer que l'ingénierie lexicale

* Ce texte constitue la version écrite d'une présentation ayant eu lieu lors d'une journée d'études organisée à Cergy-Pontoise intitulée «La lexicographie, rayonnement d'hier à demain et lexiculture» le 20 novembre 2009. Des éléments présentés ici reprennent les idées développées dans l'ouvrage *Chanson: son histoire et sa famille dans les dictionnaires de langues française*, Berlin/New York, Walter de Gruyter, 2010.

s'inscrit dans une démarche intellectuelle faite de conventions et de vide lexical (syntagme que j'emploie par analogie à vide juridique). L'ingénierie lexicale répond aux besoins des linguistes et théoriciens des sciences humaines qui créent des dénominations, étudient le sens en contextes ou les significations de dénominations préexistantes. Ainsi à partir d'une étude sur la définition prototypique de *grève* dans les dictionnaires courants actuels, Tournier (2002: 277) en vient à la conclusion suivante:

Seule une définition prototypique, établie à partir de réponses faites à un questionnaire sociolinguistique, serait vraiment synchroniste, donc structuraliste. Une véritable enquête interrogeant un panel représentatif de la population française adulte est seule légitimée à apporter ces réponses. Elle ferait peut-être apparaître qu'aujourd'hui la grève est ressentie comme un phénomène plus général et moins spécialisé qu'autrefois. Elle montrerait sans doute aussi le flou des significations qui préside à l'usage des mots.

Voilà résumé les problèmes auxquels Tournier a été confronté en faisant une analyse diachronique de certains lexèmes du champ sémantique et des co-courants de *grève*.

Devant une telle situation que ce soit l'analyse d'un phénomène social ou d'une discipline, le théoricien bénéficie souvent du travail du métalexigraphe comme le montrent les nombreux ouvrages d'introduction ou synthétique en sciences humaines, commençant par une définition très succincte de la discipline étudiée, définition souvent extraite, en langue française, du *Petit Larousse* ou du *Petit Robert*. Mais il faut constater les limites de ces définitions lorsque le théoricien en vient aux prises à des problèmes similaires à ceux que soulève Tournier. Pour cette raison j'ai créé le terme «ingénierie lexicale», et expliqué l'activité de résolution de problèmes théoriques. Ainsi, l'activité néologique propre à l'ingénierie lexicale constitue un ensemble de prises de décisions cherchant à répondre aux problèmes soulevés lors de l'enquête terminologique en diachronie.

Ingénierie terminologique

Dans la première partie de notre texte nous avons évoqué l'utilité de l'ingénierie lexicale afin de proposer des solutions aux problèmes rencontrés dans le cadre d'une activité reliée à l'usage des vocabulaires de spécialité en diachronie. La création d'un vocabulaire supradisciplinaire constitue l'un des objectifs de l'ingénierie terminologique. Dans ce cadre, elle peut s'avérer une solution socioterminologique en satisfaisant les impératifs nominatifs et conceptuels d'une communauté linguistique et culturelle. En partant du concept d'ingénierie lexicale j'en ai déduit celui d'ingénierie terminologique. L'ingénierie terminologique consiste à étudier globalement un champ disciplinaire afin d'y proposer des solutions empiriques, par exemple la création néologique ou la création d'archi-concepts opératoires. Sur le plan cognitif, l'ingénierie terminologi-

que se décline en un va-et-vient épistémologique, une négociation sémantique qui va de l'onomasiologie à la sémasiologie, de l'induction et de la déduction. Traugott (1998: folio 5) affirme à ce sujet: «Ultimately of course, semasiology presupposes onomasiology, and onomasiology presupposes semasiology, since domains have members, and over times individual form-meaning pairs come to be used to express those domains [...].» Traditionnellement l'activité du terminologue consiste à l'élaboration du fichier de dépouillement par le repérage des unités terminologiques ce qui constitue en quelque sorte une activité de cartographie *pré-sémasiologique* d'un champ disciplinaire. Les données métaterminologiques sont extraites des monographies, dictionnaires de spécialités et servent à la réflexion sémasiologique lors de l'enquête terminologique.

Même si en apparence, la résolution d'un problème de dénomination cherche la simplification, l'ingénieur terminologique cherche plutôt à juger des *imperfections théoriques*, selon le concept de Shapere (1977), qui se manifestent entre autres, lorsque la théorie proposée se révèle incomplète relativement au domaine visé, ou ne constitue qu'une simplification. Les procédés de traitement des néonymes ont été largement commentés dans les travaux de terminologues comme Loïc Depecker, Teresa Cabré, Daniel Gouadec: les emprunts directs: emprunt intégral (importation de la forme et du sens sans modification significative), emprunt sémantique, emprunt aménagé (aménagement de la forme, et éventuellement du sens), les emprunts par traduction: calque (traduction littérale du terme étranger), calque transpositionnel (transposition non littérale), synthèse néologique (reconceptualisation) (voir Depecker 2001: 403), inversion syntaxique des éléments de la langue anglaise, etc.

Parmi les imperfections théoriques qui sont notamment liés à l'étude d'un champ disciplinaire en diachronie mentionnons la synonymie. Ainsi, le phénomène de synonymie appelle celui de *coexistence terminologique*, que nous avons créé pour désigner le fait qu'un signifié identique est nommé par des dénominations différentes selon les perspectives d'études et les périodes ou encore que des référents différents sont lexicalisés de manière similaire en un même moment donné. Les dénominations, bien que différentes, ont des sens apparentés mais non équivalents. Il s'agit donc d'un concept proche de celui de *synonymie* propre à la sémantique et de *terme associé* employé en indexation¹ mais différent en cela qu'il ajoute une dimension diachronique au phénomène de similitude sémique de même que l'idée de point de vue dénominatif divergent. *Chansonnier* et *auteur-compositeur-interprète* peuvent être considérés comme des phénomènes de *coexistence terminologique*. Ce concept désigne aussi le fait que des unités sont parfois en exclusion dans l'un de leurs sens mais sont synonymes dans l'autre.

Problèmes à résoudre dans le cadre de l'ingénierie lexicale

Les différents points que nous avons traités plus tôt nous permettent de soulever les problèmes inhérents à l'étude du lexique d'une discipline en diachronie:

- (1) Evolution historique des signifiés, des définitions
- (2) Co-existence lexicale
- (3) Hétérogénéité des activités d'un vocabuliste: activité lexicale, terminologique, mots concrets, sociaux.
- (4) Absence de dénominations pour décrire des phénomènes de l'activité ou de la discipline étudiée;
- (5) Etudes d'un objet multidisciplinaire: l'exemple de la chanson nécessite le recours à des archi-concepts.

Analyser l'évolution historique d'une activité sociale permet de comprendre les lacunes de la description lexicographique du lexique en mettant en perspective l'étude philologique des formes et des significations. Mais outre le flou des significations que l'auteur évoque avec raison, c'est aussi l'absence de dénominations de certains phénomènes qui devraient nous intéresser en tant qu'ingénieurs lexicaux. Les unités lexicales désignant des choses concrètes n'ont évidemment pas la même fluctuation que les dénominations de faits sociaux, esthétiques et sont sûrement plus proches dans leur fonctionnement logico-sémantique des termes que des lexèmes. C'est en somme là un ensemble de problèmes auxquels nous sommes confrontés en étudiant un champ lexical ou sémantique en diachronie.

Afin de résoudre les problèmes qui se présentent dans l'étude d'une activité, d'un champ scientifique, d'une discipline, nous proposons la cartographie sémantique: il faut d'abord constituer un corpus représentatif des items lexicaux à étudier. En somme il faut étudier le vocabulaire d'un champ disciplinaire au sein d'un corpus textuel important non pas tant pour réaliser à la manière de Tournier un lexicogramme, que de dégager un réseau lexical pour cerner les vides lexicaux.

- (1) Relever des imperfections théoriques par l'étude onomasiologique et sémasiologique, le repérage des coexistents lexicaux.
- (2) Procéder à un éventuel recyclage terminologique par la néonymie ou la néologie.

Les premières étapes de l'étude concernent d'une part l'étude du sens (en contexte) et de la signification (hors contexte) et d'autre part la contribution néologique par la formulation d'un lexique *supradisciplinaire* selon le néologisme construit par Wijnands (1997: 139) dans la mesure «où il comprend un vocabulaire qui s'ajoute aux vocabulaires de chacune des disciplines participantes». Cette approche cherche à favoriser l'univocité référentielle tout en restant conscient des limites pragmatiques de l'objet d'étude et sans pour autant adopter l'univocité entre le terme et le concept propre à la pensée logique émanant de la terminologie.

Le problème de la référence dans le vocabulaire chansonnier: un flou incontournable?

Il faut être conscient des limites que pose la création néologique dans le champ des études littéraires. En s'intéressant à des objets-chansons caractérisés par (selon Arrivé, Gadet en Galmiche 1986 cité par Beaumont-James 1994: 10) un «système de significations qui se manifestent simultanément par des signifiants de divers types» (rythme, mélodie, parole, image, voix, accompagnements, bruits), par des procédés de diffusion divers, le caractère subjectif est d'autant plus flagrant qu'il ne s'agit là que de constructions conceptuelles qui, en ce sens, évoquent le processus de formation des concepts propres aux sciences humaines et sociales. Si les faits historiques et sociaux considérés en pure synchronie sont immuables, en revanche certains aspects impliqués dans l'étude d'un vocabulaire, même en synchronie, ne le sont pas. Ainsi le métalangage, les conditions d'étude de ces objets, donc l'aspect épistémologique et sémiotique des phénomènes chansonniers, vont influencer sur la description lexicographique des unités lexicales de la famille morphologique de *chanson*. C'est aussi la littéarité d'une forme aussi controversée que la chanson qui est sujette aux changements. Budin (1991: 338) précise à cet effet: «Since most of these objects described above constantly change (especially in sociology, political science etc., but certainly not in historical disciplines), empirical and — as a consequence — also theoretical concepts are modified in order to represent these objects adequately. Therefore, a certain degree of inherent fuzziness cannot be prevented».²

Néologie lexicale du phénomène chansonnier

Afin d'illustrer les propos de Budin, restons dans le cadre de la *poésie vocale*. Au XIXe siècle, on n'assimilait pas forcément *chanson* et *performance vocale*. Pour cette raison, l'emploi du syntagme *poème chanté* ou *poème mis en musique* et *poème mis en chanson* a été adopté. La présence du substantif *poésie* sert ici à mentionner qu'il s'agit d'un texte. *Poème chanté* ou encore *poésie chantée* rappelle donc que la vocalisation reste virtuelle d'une part et que, d'autre part, il s'agit d'un texte à auteur connu que l'on met ensuite en musique. Boiste (1834) l'exprime explicitement à l'entrée «chanson»: «Pièce de vers par couplets, que l'on peut chanter». Comme cette dénomination recouvre un ensemble de phénomènes de diffusion d'objet-chansons relativement hétéroclite, on a préféré *chanson signée sur timbre*, s'il s'agit d'une mélodie de la tradition orale, d'un *poème mis en musique*, ou s'il s'agit d'un compositeur qui met en musique le texte d'un poète.

Cette dimension aléatoire de la vocalisation de la chanson correspond aussi au sémème /recueil de poésie/ acception généralement consignée dans les dictionnaires de langues générale au XIXe siècle. Ces observations sur la nature fluctuante des sémèmes et forcément des objets esthétiques nous font adopter les vues de Traugott (1997: 7) pour qui la *subjectivation* est un phéno-

mène diachronique: «It may result synchronically in layerings of less or more subjective meanings of the same lexical item or construction, but there is no necessary isomorphism between the diachronic and synchronic variation».³ Les données sont linguistiquement hautement construites, mentionne-t-elle, et peuvent voir des sens anciens et nouveaux coexister synchroniquement.

Cette approche conceptologique met en valeur le caractère subjectif de la stabilité des concepts mais aussi de leur élaboration. Elle fait bien voir en outre que l'idéal lexicographique d'une représentation mentale abstraite et générale n'échappe pas au contexte institutionnel et culturel dans lequel il est élaboré.

L'ensemble des procédés d'*ingénierie terminologique* vise à restaurer la puissance polysémique, à tenir compte de la variation diachronique et à étudier les éventuelles absences de dénominations dans un champ terminologique. L'ingénierie terminologique et lexicale est un procédé métaterminologique traduisant les différents processus impliqués dans la création lexicale et dans le cycle de vie terminologique, que ce soit la *néologisation* ou la normalisation (ou de *normaison*, concept que nous devons à Guespin 1975). L'ingénierie terminologique consiste donc à proposer des éléments de solutions en réponse au vide terminologique et aux ambiguïtés pragmatiques liées aux changements de sens en diachronie et à la variabilité des contenus conceptuels sur le plan diachronique (ainsi *l'acquis de l'Union* (Traité de Nice) inclut *l'acquis communautaire* (Traité d'Amsterdam)). L'existence du syntagme *seconde génération* pour évoquer les enfants d'immigrants traduit la nécessité historique de la distinguer de la génération précédente. Il s'agit là d'un exemple d'ingénierie lexicale parmi tant d'autres dont la valeur terminologique pourrait aussi être contesté bien que le terme figure dans certains dictionnaires spécialisés (Bolaffi et al. 2003: 199). L'ingénierie terminologique s'intéresse en ce sens à la valorisation sociale de certains usages au détriment de certains autres. Elle peut être le fait d'un essayiste qui crée son vocabulaire opératoire comme elle peut être le fait d'organismes officiels qui proposent, recommandent des néonymes à une communauté intellectuelle.

L'usage d'un vocabulaire empirique

Dans le cadre de mes travaux de recherches depuis dix ans j'ai notamment divisé l'objet-chanson en deux grandes familles (la chanson de tradition orale et la chanson signée) et j'ai aussi introduit quelques dénominations qui répondent à la difficulté de cerner ce qu'est une «chanson» de ce qui n'en est pas une. Si l'on considère avec Wittgenstein ([1922], 1966: 29) que «le monde est l'ensemble des faits, non pas des choses», nous déduisons celui de *fait chansonnier* qu'utilise Marcadet (1999) sans doute inspiré du concept de Green (1998) issu de l'anthropologie française *fait musical*. Nous préférons l'emploi de *phénomène chansonnier* à cette première dénomination qui nous semble objectiviste, voire positiviste. D'abord comment cerner les contours conceptuels d'un genre? Dans la mesure où le mot *chanson* regroupe des esthétiques⁴ des objets-chansons

divers en plus d'être étudié selon de multiples points de vue, il convient en ce sens d'utiliser *phénomène chansonnier*. Le phénomène chansonnier signifie l'ensemble des pratiques, manifestations ou traditions, et supports diffusant et médiatisant l'objet-chanson, d'une part, et, d'autre part, l'ensemble des recherches et publications sur la chanson ou qui influencent son statut dans les champs culturel, économique et symbolique. Cela inclut toutes les monographies ou articles concernant la chanson (biographie, étude, essais, etc.). Décrire un phénomène chansonnier, c'est donc parler d'une multitude d'objets-chansons possédant certaines caractéristiques communes sur le plan du processus créatif et des canaux de diffusion (Wierzbicka 1985: 191).⁵ Le concept de phénomène chansonnier renvoie donc plus à un construit social, historique et historiographique qu'à l'objet chanson lui-même.

Si j'ai préféré phénomène chansonnier au concept déjà existant de *fait chansonnier* (voir entre autres Marcadet 1999: 291), qui renvoie à l'origine des chansons, aux performances et aux impacts qu'elles suscitent, c'est à cause d'une vision propre à la théorie de la connaissance qui fait du phénomène, au sens kantien du terme, non une chose en soi mais une donnée de l'expérience. En ce sens, relève de *chanson* ce qui est considéré comme tel par la conscience. Dans ce cas il devient *objet-chanson*. Par *objet-chanson*, j'entends le résultat final du procédé d'interprétation du phénomène chansonnier et l'objet généralement reconnu comme une chanson (celle décrite par le sémème: /pièce de vers chanté sur un air/), bref l'objet de la performance. L'objet-chanson constitue un sous-ensemble des phénomènes chansonniers et sert à alimenter l'activité chansonnaire, laquelle est tout aussi bien relative à l'activité de publication de recueils, qui a retenu l'intérêt des médiévistes, qu'à celle des médiateurs qui «colportent» des chansons, laquelle a surtout retenu l'attention des folkloristes.

Recyclage terminologique

J'ai évoqué plus tôt dans les solutions que propose l'ingénieur aux imperfections théoriques le recyclage terminologique. Contrairement au marquage en lexicologie qui fait cohabiter les marques d'usages obsolètes ou anciens et les marques de nouveautés, de celui des toplectismes où prédomine le marquage des usages archaïques au détriment des usages nouveaux, celui des termes revêt plutôt le caractère d'*institutionnalisation du signe linguistique* (selon la théorie d'Alinei 1997) et diverge du parcours de la consécration lexicographique dont le point de départ méthodologique, du moins en théorie, est le texte. En observant les trois phases du cycle de vie terminologique (sa naissance, sa vie et sa mort), on prend conscience de la vitalité d'un emploi. Les pratiques lexicologiques et terminologiques diffèrent sur ce point. On peut en effet remarquer que le fait de consigner une unité relève d'un choix dans la description d'un certain usage (lexicologie) ou d'une stratégie d'implantation d'un terme récemment proposé par un organisme officiel (terminologie). La trajectoire métalinguistique se prête ainsi à une lecture thématique, organisée sur un

champ lexical et sémantique comportant un effort d'historicisation, retraçant le cycle de vie des termes et des notions (voir Collinot et Mazière 1997: 1). Ce sont là des preuves de la vitalité d'un mot puisque l'inscription d'un lemme dans la nomenclature pour la première fois passe d'un usage réduit, dans l'absolu idiolectal, à un signe public (ou plutôt *publicisé*). L'étape ultime de changement consiste à adopter un néologisme et pourrait se nommer *riciclaggio dei segni preesistenti* suivant le concept d'Alinei (1997: 16). Ce «recyclage terminologique» peut être notamment le fait d'une action d'un *médiateur métaterminologique* (comme les commissions de terminologie) qui exercent, depuis les années 1970, le rôle de critique vis-à-vis des emplois en vue de suggérer des solutions néonymiques pour la description d'un champ d'activité supradisciplinaire.

Conclusion

La théorie générale de la terminologie et les contraintes de la mouvance référentielle sont l'objet de cet article.

Bien qu'il ait été ici procédé pour une partie de notre exposé à un travail d'onomasiologie où normalement les concepts sont nécessairement construits en vue d'édifier un arbre conceptuel, un champ notionnel bien délimité, on montre quelques réserves envers ce modèle traditionnel logico-déductif de formation des concepts.⁶ Nous sommes d'avis que l'étude d'un vocabulaire d'une part, et des phénomènes chansonniers d'autre part, implique des procédés d'abstraction qui ne s'excluent pas l'un l'autre. Le vocabulaire se place du côté du signifiant, les phénomènes chansonniers du côté du référent et les significations du côté des constructions conceptuelles.

Le processus d'*ingénierie lexicale* induit par celui d'un lexique explicatif supradisciplinaire nous a fait percevoir que le flou du vocabulaire chansonnier n'est pas uniquement dû à une méconnaissance des processus de composition, des registres, des caractéristiques profondes qui permettent une meilleure identification des genres chansonniers, mais aussi et surtout à ce que l'objet-chanson est un objet bâtard.

En restant conscient de cette nature fluctuante, éphémère des concepts et dénominations propres aux sciences humaines, nous avons proposé un lexique explicatif. Cette approche conceptologique met en valeur le caractère subjectif de la stabilité des concepts mais aussi de leur élaboration. Elle fait bien voir en outre que l'idéal lexicographique d'une représentation mentale abstraite et générale n'échappe pas au contexte institutionnel et culturel dans lequel il est élaboré. Il n'échappe pas non plus aux limites des corpus aussi bien dans la tradition d'érudition des lexicographes traditionnels que dans les corpus électroniques, qui, en langue française du moins, sont encore peu représentatifs de la fréquence d'usage de certaines unités complexes comme *chanson populaire*.

Dans la mesure où nous considérons globalement l'ingénierie lexicale comme la science de régulation, de recyclage des métatermes mais aussi plus généralement la science de la résolution des problèmes lexicaux et terminologique,

rien ne nous empêche de considérer certaines pratiques d'auteurs au sein de l'ingénierie lexicale, comme c'est le cas d'Hubert Aquin chez qui l'ont trouvé des lexiques qui sont des dossiers préparatoires ayant servi à l'élaboration de certains romans, certains étant, selon l'expression même de l'auteurs des «mots générateurs» (Dupuis 2005: 49).

Notes

1. Nous renvoyons à la terminologie du guide RAMEAU (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié 1995).
2. Traduction libre: «Dès lors que ces objets décrits [...] changent constamment (spécialement en sociologie, science politique, etc. mais certainement pas dans les disciplines historiques) les concepts empiriques, et par conséquent théoriques, sont modifiés dans le but de représenter ces objets adéquatement. Un certain degré de flou intrinsèque ne peut être alors évité».
3. Traduction libre: «Il peut résulter synchroniquement dans la superposition d'un ou de plusieurs sens subjectifs de la même unité ou construction lexicale, mais il n'y a pas d'isomorphisme entre la variation diachronique et synchronique».
4. Notons les affirmations critiquées de Theodor Adorno en ce domaine parce qu'il prétendait que les caractéristiques esthétiques de la chanson de masse constituaient la moyenne, ni plus ni moins, de l'ensemble des chansons présentées sur le marché. La chanson, pour se vendre, doit faire l'objet d'un rapport avec les goûts de la masse et devenir ainsi un objet de consommation soumis aux mêmes conditions de mises en marché que les biens capitalistes.
5. Il existe tout autant d'objets-chansons que d'occurrences de phénomènes chansonniers activant la facette /pièce de vers/. Je ne considère pas comme objet-chanson les références de /parole/, /rengaine/ qui sont les sens figurés (voir Du Chazaud 1991, s. v. chanson). Pour une liste exhaustive de synonymes, on se reportera au dictionnaire signé par l'équipe de Sabine Ploux, Jean-Luc Manguin, et Bernard Victorri qui ont constitué une base de départ de sept dictionnaires de langue française (Bailly, Bénac, Du Chazaud, Guizot, Lafaye, Larousse et Robert). Il donne 62 synonymes de *chanson*. Disponible à l'adresse <http://elsap1.unicaen.fr/dicosyn.html>. Juin 2003.
6. Notre réflexion est conduite en partie par l'ouvrage de Bolton (1977).

Bibliographie

- Alinei, Mario.** 1997. Principi di teoria motivazionale (iconimia) e di lessicologia, motivazionale (iconomastica). Mucciante, Luisa et Tullio Telmon (Éds.). *Lessicologia e Lessicografia. Atti del Convegno della Società Italiana di Glottologia. Chieti-Pescara 12-14 ottobre 1995*: 9-36. Rome: Il Calamo.
- Arrivé, Michel, Françoise Gadet et Michel Galmiche.** 1986. *La grammaire d'aujourd'hui*. Paris: Flammarion.
- Beaumont-James, Colette.** 1995. Analyse sémantique du mot *chanson*. *Cahiers de lexicologie* 67(2): 163-192.
- Bolton, Neil.** 1977. *Concept Formation*. Oxford: Pergamon.
- Budin, Gerhard.** 1988. The Application of Terminology-based Knowledge Data Bases in the Humanities and the Social Sciences and Its Impact on Research Methods. Best, Heinrich,

- Ekkehard Mochmann et Manfred Thaller (Éds.). 1991. *Computers in the Humanities and the Social Sciences*: 337-342. München/Londres/New York/Paris: Saur.
- Collinot, André et Francine Mazière.** 1997. *Un prêt à parler: le dictionnaire*. Paris: PUF.
- Depecker, Loïc.** 2001. *L'invention de la langue: Le choix des mots nouveaux*. Paris: Armand Colin, Larousse.
- De Surmont, Jean-Nicolas.** 2010. *Chanson: son histoire et sa famille dans les dictionnaires de langue française*. Berlin/New York: Walter de Gruyter.
- De Surmont, Jean-Nicolas.** 2005. L'ingénierie terminologique comme contribution à la constitution d'un vocabulaire supradisciplinaire. *Le langage et l'homme* 40(1): 123-136.
- Dupuis, Gilles.** 2005. Du mot juste au mot rare. Dictionnaires et glossaires dans l'œuvre d'Hubert Aquin et Jacques Poulin. Acerenza, Gerardo (Dir.). 2005. *Dictionnaires français et littératures québécoise et canadienne-française*: 35-50. Ottawa: Éditions David.
- Green, Anne-Marie.** 1998. *Musicien de métro. Approches vivantes urbaines*. Paris/Montréal: L'Harmattan.
- Guespin, Louis.** 1975. Normaliser ou standardiser? *Le langage et l'homme* 28(4): 213-222.
- Guespin, Louis.** 1985. Matériaux pour une glottopolitique. *Problèmes de glottopolitique, Cahiers de Linguistique Sociale* 7: 13-32.
- Marcadet, Christian.** 1999. La matrice des faits-chansons: le cas Yves Montand. Quénart, Jean (sous la direction de). 1999. *Le Chant, acteur de l'histoire*: 291-306. Rennes: Presses Universitaires de Rennes.
- Shapere, Dudley.** 1977. Scientific Theories and their Domains. Suppe, Frederick (Éd.). 1977. *The Structure of Scientific Theories*: 518-599. Deuxième édition. Urbana: University of Illinois Press.
- Tournier, Maurice.** 2002. *Des mots sur la grève. Propos d'étymologie sociale 1*. Lyon: ENS Éditions.
- Traugott, Elizabeth Closs.** 1997. The Rhetoric of Counter-expectation in Semantic Change: A Study in Subjectification. Blank, Andreas et Peter Koch (Éds.). 1999. *Historical Semantics and Cognition*: 177-196. Berlin/New York: Mouton de Gruyter.
- Wierzbicka, Anna.** 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma Publishers.
- Wijnands, Paul.** 1997. L'interdisciplinarité du discours sur l'identité culturelle: les mots d'auteur. De Schaetzen, C. (Éd.). 1997. *Terminologie et interdisciplinarité. Actes du colloque organisé en avril 1996 par le Centre de terminologie de Bruxelles (Institut Libre Marie Haps) et l'Association européenne des Professeurs de Langue vivantes (AEPLV)*: 137-146. Louvain-la-Neuve: Peeters.
- Wittgenstein, Ludwig.** 1961. *Tractatus logico-philosophicus, suivi des investigations philosophiques*. Paris: Gallimard. (L'édition d'origine est celle de Londres: Kegan Paul, 1922.)

Dictionnaires

- Boiste, Pierre Claude Victoire.** 1834. *Dictionnaire universel de la langue française, avec le latin et les étymologies [...]*. Huitième édition [revue par C. Nodier]. Paris: Lecointe et Pougin, Quai des Augustines. S.v. «Chanson».
- Bolaffi, Guido, Raffaele Bracalente, Peter Braham et Sandro Gindro.** 2003. *Dictionary of Race, Ethnicity & Culture*. Londres: Sage Publications.
- Du Chazaud, Henri Bertaud.** 1991. *Dictionnaire de synonymes et contraires*. Paris: Le Robert.
- [En Coll.]. 1999. [PLI] *Le petit Larousse illustré*. Paris: Larousse.
- Rey-Debove, Josette et Alain Rey (texte remanié et amplifié sous la dir. de).** 1993. [NPR1] *Le Nouveau Petit Robert, dictionnaire alphabétique et analogique de la langue française*. Paris: Dictionnaires Le Robert.

Le dictionnaire comme outil d'enseignement des langues au Gabon

Guy-Modeste Ekwa Ebanéga, *Département des Sciences du Langage,
Université Omar Bongo, Libreville, Gabon (guymodeste_e@yahoo.fr)*

Résumé: Le but de cet article est de montrer dans quelle mesure le dictionnaire peut être un outil d'enseignement des langues gabonaises. Dans un pays, comme le Gabon, caractérisé par une pluralité de langues locales et étrangères, le dictionnaire sera non seulement un ouvrage de référence contenant l'ensemble des mots d'une langue et fournissant pour chacun des données phonétiques, morphologiques, sémantiques, étymologiques etc. Il sera également un outil didactique permettant aux étudiants et aux élèves d'acquérir une compétence nécessaire de leurs langues. Il sera encore un moyen efficace qui va permettre aux usagers de maîtriser leurs langues et dépasser les frontières linguistiques qui existent entre les langues.

Mots clés: DICTIONNAIRE, LANGUES GABONAISES, ENSEIGNEMENT, UNITÉS LEXICOGRAPHIQUES, STRUCTURE DU DICTIONNAIRE, PHONÉTIQUE, MORPHOLOGIE, SÉMANTIQUE, VOCABULAIRE, ÉTYMOLOGIE

Abstract: *The Dictionary as Tool for Teaching the Gabonese Languages.*

The aim of this article is to show to what extent the dictionary can be a tool for teaching the Gabonese languages. In a country like Gabon, characterized by a multiplicity of local and foreign languages, the dictionary will not only be a reference work containing a collection of the words of a language and providing for each its phonetic, morphologic, semantic and etymologic etc. data. It will also be a didactic tool allowing students and pupils to acquire the necessary competence in their languages. It will further be an efficient means which will allow users to master their languages and to overcome the linguistic boundaries existing between the languages.

Keywords: DICTIONARY, GABONESE LANGUAGES, TEACHING, LEXICOGRAPHIC UNITS, DICTIONARY STRUCTURE, PHONETICS, MORPHOLOGY, SEMANTICS, VOCABULARY, ETYMOLOGY

1. Introduction

Les dictionnaires ont été compilés et utilisés pendant plusieurs siècles (Al-Kasimi 1977; McArthur 1986). Ils sont nés d'une situation de transfert de connaissance. Le dictionnaire, comme instrument pratique, répond à une pratique de communication au niveau du décodage et il satisfait un type bien défini d'usagers: des étudiants, des élèves et tous ceux qui se trouvent en situation

d'apprentissage d'une langue étrangère (Gouws 2000). Les dictionnaires sont des livres de mots (au sens premier, «Wörterbuch» en allemand). Ces mots traduisent les pensées d'un groupe ou d'une nation qui se transmettent de personne en personne et de génération en génération. À ce titre, le dictionnaire est un patrimoine linguistique et culturel à léguer aux générations futures.

Trois périodes caractérisent la lexicographie gabonaise (Ekwa Ebanéga et Tomba Moussavou 2008). Pendant la première et la seconde période, les missionnaires sont les premiers à s'intéresser aux langues gabonaises et compileront les dictionnaires du gabonais. Ces dictionnaires sont tous des dictionnaires bilingues à l'usage du public francophone. La troisième période est celle des lexicographes nationaux qui ont reçu une formation de base en lexicographie et en planification lexicographique pour les langues gabonaises.

Le but de cet article est de montrer dans quelle mesure le dictionnaire peut être considéré comme un outil didactique important pour l'éducation des apprenants au Gabon. Notre premier but sera la mise en place d'une politique lexicographique (2), suivi d'une analyse des problèmes lexicographiques liés à l'éducation au Gabon (3), d'une interrogation sur des dictionnaires se prêtant bien au mode d'instruction (4), d'une valorisation du dictionnaire dans le système éducatif gabonais (5), d'une réflexion sur l'établissement des unités lexicographiques du Gabon (6), d'une identification du public visé: les enseignants, étudiants et élèves (7), d'une insistance sur l'importance d'utiliser le dictionnaire comme outil d'enseignement des langues (8) et d'une exposition des aspects de la lexicographie (9).

2. La mise en place d'une politique lexicographique

Le Gabon est un pays multilingue. Kwenzi-Mikala (1998: 217) dénombre 64 parlars, parmi lesquels les langues bantoues et pygmées. Sur le plan de l'éducation, la catégorisation retenue distingue la langue officielle, c'est-à-dire le français, et les langues internationales (anglais, espagnol, allemand, portugais, italien et arabe), enseignées aux niveaux secondaire et supérieur. Quelques langues gabonaises telles que le fang, le yipunu, le yinzebi et l'omyéne sont enseignées dans certains établissements secondaires catholiques de Libreville, notamment à l'Institut de l'Immaculée Conception, à Quaben, à Bessieux et à Sainte Marie.

Depuis les États Généraux de l'Éducation Nationale 1983, le gouvernement gabonais s'engage à promouvoir les langues du pays, afin de sauvegarder les cultures gabonaises, ainsi que par l'insertion de ces langues dans l'éducation.

En 1984, le Gabon s'efforce de mettre en place une politique linguistique qui contribuerait au développement de ses langues nationales, aux côtés de la langue française. Au cours des dernières années, la linguistique s'est développée au Gabon, et nous pouvons observer l'apparition des travaux universitaires, qu'il s'agisse des outils pédagogiques et didactiques (livres de lecture ou

d'écriture, contes, etc.), ou de lexiques pour l'enseignement des langues gabonaises. À l'Université Omar Bongo, des recherches linguistiques se créent pour «penser les langues gabonaises». Nous pouvons citer entre autres le Groupe de Recherche en Langues et Cultures Orales (GRELACO), le Centre de Recherche en Langues et Linguistique (CRELL), le Laboratoire Universitaire de Tradition Orale (LUTO), et le Laboratoire des Sciences de l'Homme et de la Dynamique du Langage (LASCIDYL). Ces structures de recherche sont destinées à sauvegarder et à promouvoir les langues locales pour l'élaboration des projets de recherche.

Un accord de coopération liant le Groupe de Recherches en Langues et Cultures Orales (GRELACO) de l'Université Omar Bongo et le Bureau du *Woordeboek van die Afrikaanse Taal* (WAT) fut signé le 6 décembre 1999. Également en 1999, deux étudiants gabonais sont venus au Bureau du WAT pour une formation continue et pour leur étude doctorale en lexicographie à l'Université de Stellenbosch (République d'Afrique du Sud) sous la direction du professeur R.H. Gouws. En 2000, trois autres étudiants sont venus en Afrique du Sud et en 2001, ils ont été suivis par cinq autres. Le Bureau du WAT leur a donné une formation pilote en lexicographie générale, en lexicographie informatique et en planification des projets lexicographiques. Cinq étudiants ont soutenu leur thèse de doctorat en lexicographie.

Trois ans plus tard s'est tenue à Libreville du 18 au 23 juillet 2004, la Neuvième Conférence Internationale de l'Association Africaine pour la Lexicographie (AFRILEX). Le sujet de cette conférence était la problématique du dictionnaire, le développement des langues et l'organisation des centres lexicographiques. Son objet était la recherche des zones de convergences entre le dictionnaire et la langue, afin de proposer une stratégie dynamique pour le développement de la lexicographie gabonaise. Le dictionnaire est un moyen qui rend accessibles les connaissances du monde contemporain aux élèves et aux étudiants qui sont appelés à le consulter pour la pluralité des langues et des cultures du Gabon.

Depuis 2004, la lexicographie est enseignée au Département des Sciences du Langage à l'Université Omar Bongo. On y dénombre, en ce moment, quatre lexicographes-enseignants. Pour la période 2006–2008, le CENAREST (Centre National pour la Recherche Scientifique et Technologique) a, à son tour, recruté deux lexicographes-chercheurs.

3. Les problèmes lexicographiques liés à l'éducation au Gabon

Il semble toutefois que les dictionnaires existant en langues gabonaises sont très peu utilisés au Gabon. Ils ne font pas partie des ouvrages courants à consulter, et ils sont rangés dans la catégorie des ouvrages occasionnels. Les dictionnaires en langues gabonaises ne se trouvent pas dans les librairies, ni sur les rayons des bibliothèques générales, ni dans les établissements scolaires et universitaires.

L'observation de McKean à propos des jeunes enseignants d'anglais est aussi valable pour les jeunes enseignants gabonais pour qui leur cours n'implique aucune information sur les dictionnaires et rien sur l'instruction des connaissances dictionnairiques. Il faut noter que bon nombre d'enseignements sont dispensés en langues étrangères (français, anglais, allemand, espagnol, etc.) et les étudiants sont familiers avec ces langues. Cela n'est pas le cas des langues du Gabon où nous notons une absence non seulement de l'enseignement des langues nationales mais aussi d'un enseignement des connaissances phonétiques, phonologiques, grammaticales, morphologiques, syntaxiques, sémantiques et pragmatiques de ces langues.

Le gouvernement gabonais ne fournit pas de dictionnaires ou autres ouvrages lexicographiques (qui existent en langues gabonaises) aux écoles; il ne subventionne même pas l'acquisition de ces matériaux pour stimuler les études appropriées. La question que nous pouvons poser consiste à savoir comment les connaissances dictionnairiques peuvent être appliquées et encouragées dans un environnement où elles ne font pas partie du programme d'étude offert dans les établissements, les universités et les cours offerts dans les écoles?

Certaines écoles n'ont pas de bibliothèques et n'encouragent pas la consultation des dictionnaires pour aider l'usager à acquérir les techniques et connaissances dictionnairiques. Ou sinon, la question existe combien de dictionnaires et de types de dictionnaires devraient être maintenus dans les bibliothèques envisagées?

Bon nombre d'enseignants de langues n'ont pas de dictionnaires en classe. Les dictionnaires qu'on trouve dans les salles de classes sont, le cas échéant, destinés à l'usage personnel de l'enseignant, et non pas pour encourager les apprenants à les utiliser.

Certains parents préfèrent acheter des manuels bibliques transcrits en langues (catéchismes, cantiques, évangiles, etc.), ou d'outils pédagogiques et didactiques (livres de lecture ou d'écriture, syllabaires, etc.) qui se trouvent disponibles sur le marché. Cela n'est pas le cas des dictionnaires en langues gabonaises qui deviennent de plus en plus introuvables. Les dictionnaires anglais, français, espagnols et portugais font partie des ouvrages didactiques, or pas les dictionnaires en langues gabonaises.

Les enquêtes menées par Ekwa Ebanéga et Tomba Moussavou (2008) ont montré qu'à la question quel(s) type(s) de dictionnaire(s) existent en langues gabonaises, 36% des étudiants ont répondu en faveur du dictionnaire général, 10% pour le dictionnaire spécialisé, 46% pour le dictionnaire bilingue, et 46% pour l'encyclopédie. Ekwa Ebanéga et Tomba Moussavou (2008) ont montré que les différentes réponses données par les étudiants sont la preuve qu'ils ne savent pas quel type de dictionnaire existe en langues gabonaises. Les dictionnaires existants en langues gabonaises ne sont pas connus par le public, et particulièrement par les étudiants et les élèves. Ekwa Ebanéga et Tomba Moussavou (2008) ont suggéré qu'un cours sur *l'histoire de la lexicographie gabonaise*, dont le but est de promouvoir les dictionnaires existants, devrait être enseigné

dans les écoles, les collèges, les lycées et les universités.

Les recherches faites par Kipfer (1987: 45) ont montré que les étudiants sont réticents à consulter les dictionnaires pour la simple raison qu'ils n'ont pas assez de compétence pour trouver les réponses aux innombrables problèmes auxquels leur langue fait face. Selon lui, la plupart des étudiants qui n'ont pas reçu des enseignements sur le dictionnaire consulte le dictionnaire pour trouver l'orthographe et la signification/l'équivalent du mot. Les informations telles que les exemples et les collocations sont rarement utilisées. L'observation de Kipfer (1987) est similaire aux résultats trouvés par Ekwa Ebanéga et Tomba Moussavou (2008) où 74% des étudiants gabonais consultent les dictionnaires pour trouver la signification/l'équivalent du mot et 58% consultent le dictionnaire pour trouver l'orthographe du mot.

Les linguistes et les lexicographes ne sont pas nombreux au Gabon. On y dénombre environ vingt linguistes et dix lexicographes.

Certaines personnes pensent qu'on n'a pas besoin de faire de longues études pour produire des dictionnaires. Il suffit simplement de recueillir les mots.

Bon nombre de personnes pensent que la promotion et le développement des langues sont l'affaire des linguistes seuls, ignorant qu'elles sont les garants de son usage et les acteurs de son évolution.

Certaines personnes pensent que le lexicographe n'est pas fait pour enseigner. Cependant, elles ignorent que, pour être lexicographe, il faut avoir suivi, au préalable, une formation lexicographique. Les lexicographes gabonais ont reçu cette formation qui donne accès au doctorat (cf. Gouws 2001: 98). L'auteur continue en affirmant que l'importance et la valeur académique ne doivent jamais être sous-estimées.

4. Les dictionnaires et les langues nationales

La question du pour ou contre du dictionnaire comme outil d'enseignement dans le système éducatif gabonais a été longuement discutée par Ekwa Ebanéga (2007), Tomba Moussavou (2007), Nyangone Assam (2001) et Mabika Mbokou (2001), ces auteurs soulignant l'importance du dictionnaire dans l'enseignement des langues gabonaises. Ceci nous amène à nous interroger sur les langues qui seront utilisées comme mode d'instruction, c'est-à-dire les langues d'enseignement et d'apprentissage dans le système éducatif gabonais. Moyo (2002: 151) envisage l'activité de l'enseignement des langues de deux manières:

- (a) Le mode d'instruction, c'est-à-dire la langue dans laquelle communique l'apprenant et l'enseignant, est la langue par laquelle l'enseignant transmet le savoir à l'apprenant.
- (b) La matière d'apprentissage, c'est-à-dire la langue introduite dans le système éducatif pour acquisition par les apprenants, la langue est ici le savoir à transmettre.

L'observation faite par Moyo (2002) relativement à l'enseignement et à l'apprentissage des langues, nous amène à voir le statut des langues gabonaises. Kwenzi-Mikala (1998) dénombre 64 parlers, parmi lesquels les langues bantoues et pygmées. Le français jouit d'un statut particulier au Gabon, c'est la seule langue officielle, l'unique mode d'instruction du système éducatif national, des affaires économiques et commerciales, de l'administration civile et militaire (Nyangone Assam et Mavoungou 2000, Ndinga-Koumba-Binza 2005). Sur le plan de l'éducation, la catégorisation retenue distingue la langue officielle, c'est-à-dire le français, et les langues internationales (anglais, espagnol, allemand, portugais, italien et arabe), enseignées aux niveaux secondaire et supérieur. Quelques langues gabonaises telles que le fang, le yipunu, l'omyéné et le yinzebi sont enseignées dans certains établissements secondaires catholiques de Libreville. Nous pouvons citer entre autres l'Institut Immaculée Conception, à Quaben, à Bessieux et à Sainte Marie.

Ainsi, pour le cas du Gabon, nous pensons que le dictionnaire comme instrument d'enseignement des langues gabonaises doit se faire à partir du primaire voire dès la maternelle. Une enquête menée par Ekwa Ebanéga et Tomba Moussavou (2008) est présentée ici pour confirmer ce fait.

Par ailleurs nous pensons également qu'un outil d'enseignement comme le dictionnaire doit aller de pair avec l'intégration des langues gabonaises dans le système éducatif. Une enquête menée par Bokoko (2004) auprès des acteurs des langues nationales (professeurs et membres de l'administration scolaire) est une preuve manifeste pour confirmer cette assertion. Le but de l'enquête était de vérifier l'exactitude des informations sur l'intégration des langues nationales. Selon le vœu exprimé par la majorité des informateurs, les langues nationales devraient être enseignées à partir du primaire aux seules fins d'éviter une rupture brutale avec le milieu familial.

Ainsi, pour le cas du Gabon et prenant en compte les enquêtes menées par Ekwa Ebanéga et Tomba Moussavou (2008) et Bokoko (2004), un outil d'enseignement comme le dictionnaire doit aller de pair avec l'intégration des langues gabonaises. Leur intégration devrait permettre de

- préserver les racines et les cultures gabonaises,
- faciliter l'apprentissage d'autres langues,
- enrichir et améliorer l'expression,
- renforcer l'unité et l'identité nationales, et
- pallier à l'absence de communication en langues dans les foyers.

Pour permettre aux langues gabonaises d'être enseignées par le biais des dictionnaires, il faut que l'État gabonais définisse une politique claire en matière de la promotion des langues nationales. Pour ce faire, il faudrait répondre aux questions suivantes:

- Veut on introduire le dictionnaire comme moyen d'enseignement des langues au Gabon?
- Veut on utiliser le dictionnaire comme marche-pied pour faciliter l'apprentissage de nos langues?

En d'autres termes, il faut que les responsables gabonais répondent à la question suivante: dans quel but faut-il créer des dictionnaires de langues gabonaises? Des réponses objectives à ces questions permettront d'accorder une place de choix au dictionnaire. Cependant le bref aperçu sur la place du dictionnaire dans l'enseignement des langues montre que l'État gabonais n'a pas encore trouvé de réponse à ces interrogations. Une fois que les responsables du système éducatif auront trouvé une réponse à ces questions, l'État devrait s'appliquer davantage aux innovations susceptibles d'entraîner un renforcement de la politique linguistique dans le pays.

5. Valoriser le dictionnaire dans le système éducatif gabonais

La valorisation du dictionnaire dans l'enseignement des langues gabonaises n'est possible qu'avec l'organisation des conférences, des ateliers et des séminaires dans le but d'instruire les enseignants et les étudiants sur le rôle, la structure et le contenu des dictionnaires. En 1981, par exemple, le linguiste et lexicographe tchèque Ladislav Zgusta a mené un atelier de table ronde sur les dictionnaires bilingues à l'Université d'Illinois. Plus récemment, Kenneth Hill du Bureau de la Recherche Appliquée en Anthropologie, l'Université d'Arizona, a enseigné un cours de deux semaines sur la réalisation du dictionnaire à 20 étudiants dans le programme de linguistique à l'Université de Sonora. À cet effet, il a partagé avec les étudiants ce qu'il avait appris en faisant un dictionnaire hopi aux seules fins de les guider dans la réalisation des dictionnaires. Gates (1997) souligne que depuis 1979, plusieurs ateliers se sont organisés dans au moins sept pays. En 1978, Reinhard Hartmann a organisé un séminaire sur la lexicographie à l'Université d'Exeter financée par l'Association Britannique de Linguistique Appliquée. En mars 1989, le Centre de Recherche du Dictionnaire à Exeter a tenu un atelier sur la lexicographie africaine. En Afrique, il se tient chaque année une conférence sur la lexicographie africaine organisée par AFRILEX (l'Association Africaine pour la Lexicographie), et certains ateliers sont tenus en collaboration avec AFRILEX.

Les initiatives, ci-dessus, mentionnées au Gabon, auront essentiellement pour objectif la valorisation du dictionnaire dans le système éducatif. En termes plus concrets, la tenue des conférences, des séminaires et des ateliers au Gabon permettront aux enseignants et étudiants de se familiariser avec les connaissances dictionnaires. Les lexicographes gabonais pourraient, par exemple, organiser au Gabon un séminaire dont l'objet est "le dictionnaire comme instrument d'enseignement des langues". Ce séminaire aurait comme slogan "Donne moi

ton dictionnaire, j'évaluerai ta langue".¹ On pourrait expliquer ce slogan en disant qu'on ne peut pas espérer une éducation efficiente et développée sans considérer une langue. La connaissance linguistique se trouve dans le dictionnaire qui contient des données linguistiques telles que la phonétique, la phonologie, la morphologie, la syntaxe, la sémantique, d'où la nécessité de promouvoir le dictionnaire. Nul n'ignore que toute connaissance transmise dans la langue maternelle est mieux comprise que si elle était transmise dans une langue étrangère. Les dernières recherches faites par Joubier (2008)² le prouvent. On apprend mieux dans sa langue maternelle. Les enfants dont l'éducation a commencé dans leur langue maternelle prennent un meilleur départ, et réussissent mieux par la suite, que ceux dont la scolarité a débuté dans une langue autre que la leur.

6. L'établissement des unités lexicographiques

La question de l'établissement des unités lexicographiques est un facteur déterminant si l'on veut que le dictionnaire soit un instrument d'enseignement des langues gabonaises. Une unité lexicographique est une structure de recherche dont l'objectif principal est la production pratique des dictionnaires. Dans le cadre de la politique linguistique des langues africaines, le projet de type PanSALB (Pan South African Language Board)³ pour les langues sud-africaines a été établi par le gouvernement de l'Afrique du Sud en 1995. Pour le cas du Gabon, un projet de type PanSALB aurait pour résultat la préservation de la diversité linguistique et l'établissement des onze unités lexicographiques telles que les unités françaises et mazuna, myene, mekana-menaa, mekona-mangote, membe, merye, metye, membre, makena, baka, et selon les dix groupes de langues⁴ préalablement cités. Celles-ci serviraient d'interface pour la collaboration entre les lexicographes et les linguistes. En établissant les unités lexicographiques au Gabon, ceci aurait comme résultat la production des dictionnaires et la préservation de la diversité linguistique sous toutes ses formes variées. L'établissement des unités lexicographiques devrait aller de pair avec la politique linguistique du pays. Les actions en faveur de la préservation de la diversité linguistique devraient être prises par le gouvernement. Nous pourrions citer entre autres l'enseignement des langues gabonaises dans les écoles, lycées, et universités gabonais; l'utilisation des langues gabonaises dans les médias (radios, net et presse) et la reconnaissance constitutionnelle et légale aux langues gabonaises. La prise en compte de ces actions par le gouvernement gabonais serait plus que vitale si nous voulons que l'établissement des unités lexicographique soit effectif.

La création d'une unité lexicographique est de rigueur compte tenu des besoins lexicographiques des langues gabonaises (Nyangone Assam et Mavoungou 2000: 252; Nyangone Assam 2001: 188). Cette structure est en train de se mettre en place et sera fonctionnelle dans le cadre du Groupe de Recherche en Langues et Cultures Orales (GRELACO) au sein de l'Université Omar Bongo

(Emejulu et Nzang-Bié 1999). Le GRELACO œuvre dans le sens de la création des unités lexicographiques avec le concours des communautés linguistiques. Pour satisfaire les besoins en matière de dictionnaires et produits dérivés, supports indispensables pour le développement et la standardisation des langues, il est d'une nécessité absolue de créer des unités lexicographiques à travers l'ensemble du territoire gabonais, et ce en fonction des besoins spécifiques des langues et des communautés linguistiques telles que les communautés françaises et mazuna, myene, mekana-menaa, mekona-mangote, membe, merye, metye, membre, makena, baka, et suivant les dix groupes de langues. Selon Emejulu (2001: 50), il est plus pratique et économique d'établir les unités lexicographiques dans les localités où se trouvent les unités-langues répertoriées. Les unités lexicographiques seront installées auprès des académies provinciales dans d'autres provinces que la province du Haut-Ogooué dont l'unité lexicographique sera installée au sein de l'Université des Sciences et Techniques de Masuku à Franceville. L'unité lexicographique du français sera établie à Libreville, la capitale du Gabon. Pour le cas de l'unité lexicographique du français, elle prendra en compte la variété du français au Gabon (Mavoungou 2002).

Toutes ces unités lexicographiques seront animées par des agents lexicographiques formés par le GRELACO. La première tâche des unités lexicographiques sera l'identification et la planification des projets lexicographiques à court, moyen et long termes. La deuxième et principale tâche est d'assumer la responsabilité du développement des corpus de différentes langues.

7. Le public visé: enseignants, étudiants et élèves

Le dictionnaire comme outil d'enseignement des langues est destiné aux personnels enseignants et aux étudiants qui auront reçu une formation de lexicographes. Ils auront des connaissances spécifiques sur des langues et la compétence de leurs langues maternelles respectives, ils auront besoin également des qualifications générales, d'expertises lexicographiques, d'expertises métalexigraphiques, et, en d'autres termes, d'une formation appropriée (Gouws 2001: 99). Les enseignants doivent, à leur tour, transmettre leur savoir aux élèves et aux étudiants. Dolezal et McCreary (1999: IX) affirment: "Puisque nous observons le développement futur de cette discipline, les soucis mutuels du professeur, du lexicographe et de l'apprenant du dictionnaire pourraient être le centre de la lexicographie pédagogique." Les enseignants, comparés aux autres chercheurs ou lexicographes, sont les mieux placés pour évaluer les besoins des étudiants et élèves. Si les enseignants reçoivent une formation de base en lexicographie et, en retour, les transmettent aux étudiants et élèves, la culture du dictionnaire pourrait se développer.

Pour le cas du Gabon, je propose que les enseignants reçoivent une formation de base en lexicographie. Ils doivent également avoir des connaissances particulières sur les langues gabonaises. Cette formation servira d'interface pour la collaboration entre les linguistes et les lexicographes qui sont appelés à

travailler ensemble dans le but de préserver la riche variété des langues gabonaises (Ekwa Ebanéga et Tomba Moussavou 2006: 243). Les étudiants et élèves profiteront du savoir des enseignants et pourront enrichir leurs connaissances.

8. L'importance d'utiliser le dictionnaire comme outil d'enseignement des langues

Pour bien comprendre l'importance d'utiliser le dictionnaire comme outil d'enseignement des langues, les enseignants et étudiants sauront que le dictionnaire est le récipient du savoir. L'enquête menée par Ekwa Ebanéga et Tomba Moussavou (2008) auprès des étudiants gabonais à l'Université de Stellenbosch et l'Université de Technologie de la Cape Peninsula confirme que 89% des étudiants pensent qu'il est important de recevoir des connaissances linguistiques par le biais de l'outil qu'est le dictionnaire. La culture du dictionnaire pourrait se développer au Gabon, particulièrement parmi les élèves et étudiants en visant d'abord les enseignants. Cependant, les étudiants ne peuvent pas acquérir des connaissances linguistiques par eux-mêmes sans l'aide des professeurs. Le centre d'intérêt serait dédié à la formation des professeurs et enseignants. Selon Hartmann (2001: 26), les enseignants ont rarement reçu la formation requise pour juger les avantages et les compétences dictionnaires appropriées pour garantir les meilleures conditions de consultation réussie. Ils ne sont que rarement en contact avec des lexicographes ou des chercheurs de dictionnaire.

Le manque de prise en compte du dictionnaire comme outil d'enseignement des langues dans le monde en général et au Gabon en particulier est commun dans la plupart des langues du monde. Landau (2001: 26) note cela: "C'est une pitié, parce que l'habitude de consulter un dictionnaire est formée tôt dans la vie, et si les qualifications pour l'utiliser sont négligées, les étudiants ne peuvent jamais être les usagers confortables du dictionnaire". Cette observation vaut pour des usagers gabonais en général et les usagers en particulier. L'utilisation du dictionnaire doit faire partie intégrante de la formation de l'enseignant. La culture du dictionnaire doit être instaurée en passant par l'enseignant. C'est l'une des raisons pour laquelle Wiegand (1984) affirme que ce n'est pas l'autorité du dictionnaire qui est mise en doute, mais qu'il faut instaurer une culture du dictionnaire.

Par ailleurs, il faut dire que le dictionnaire est né d'une situation d'apprentissage. L'importance d'utiliser le dictionnaire comme outil d'enseignement de langues reposerait sur le fait que le dictionnaire répond à un besoin de communication aux niveaux du décodage (réception) et de l'encodage (production). Au niveau de la réception, lorsque les étudiants et les élèves ne connaissent pas le sens d'un mot, le dictionnaire fournit des données suivantes: la définition, le synonyme, l'antonyme, etc. Au niveau de la production, lorsqu'ils n'arrivent pas à produire un mot, le dictionnaire leur fournira des données comme la prononciation, la morphologie, etc.

Dans le cas de l'enseignement des langues gabonaises, le dictionnaire devrait, tout d'abord, aider les étudiants et les élèves à acquérir la compétence nécessaire dans une langue. Ensuite, ils feront appel au dictionnaire, lorsque se poseront des problèmes de communication aux niveaux de la réception et de la production.

Dès que les connaissances linguistiques seront acquises chez les enseignants, on pourrait aussi bien compléter leur formation en leur enseignant plusieurs aspects de la lexicographie. Ces aspects sont les suivants:

9. Les aspects de la lexicographie

9.1 La métalexigraphie

Les enseignants auront une formation de la théorie et la pratique des dictionnaires. Cette formation reposera sur les principes qui gouvernent les dictionnaires existants conduisant à la formulation des hypothèses orientées vers la production de bons dictionnaires. Les aspects suivants de la métalexigraphie seront pris en compte au cours de cette formation:

- la recherche sur l'utilisation des dictionnaires,
- l'histoire de la lexicographie,
- la critique des dictionnaires, et
- la théorie de la lexicographie générale.

Pour ce qui est de la théorie de la lexicographie générale, les aspects suivants seront également pris en compte (Hausmann 1986 et Wiegand 1984):

- la théorie textuelle des textes lexicographiques,
- les types de dictionnaires,
- la collection et l'assemblage des données, et
- les buts des dictionnaires.

Selon Gouws (2001: 99), la formation en lexicographie et métalexigraphie doit inclure des matières suivantes:

- différents types de dictionnaires,
- structure des dictionnaires,
- compilation d'un livre d'instruction pour le dictionnaire donné,
- différents types de lemmes,
- choix de lemmes pour un dictionnaire spécifique,

- différentes manières pour ordonner les lemmes,
- différents types de catégories de données dans un dictionnaire,
- comment écrire une définition du dictionnaire,
- comment utiliser les exemples d'illustration dans un dictionnaire,
- choix de l'équivalent correct pour la traduction, et
- comment indiquer la prononciation correcte.

Les étudiants et les élèves qui sont des lexicographes en devenir auront une connaissance générale sur les aspects de la métalexigraphie. Lorsque les problèmes d'ordre théorique se poseront, les enseignants formés pourront richement illustrer ceux-ci avec des exemples pris dans la pratique lexicographique. Les programmes à enseigner seront élaborés par des lexicographes et des enseignants en fonction des cycles et des objectifs voulus.

9.2 L'histoire de la lexicographie gabonaise

Il importe que les enseignants aient une connaissance satisfaisante sur l'origine et les débuts de la lexicographie en langues gabonaises. Toutefois, ils inculqueront aux étudiants que les premiers dictionnaires en langues gabonaises ont été compilés par les missionnaires vers les années 1800. Mihindou 2001 en rend compte au cours d'une enquête sur le nombre de dictionnaires qui existent en langues gabonaises. L'auteur dénombre une quinzaine de dictionnaires disponibles. Pour de plus amples précisions, voir Mihindou (2001: 11-12). Ces ouvrages constituent une excellente source d'informations sur l'histoire de la lexicographie au Gabon. Ils pourraient servir de départ à la constitution d'une base de données en vue de la compilation des dictionnaires dans les langues gabonaises. Les enseignants montreront aux étudiants les avantages et les limites des dictionnaires existants. Il importe de garder en mémoire que les premiers dictionnaires compilés en langues gabonaises avaient pour rôle majeur de servir de support pédagogique à la lecture des langues gabonaises. Ils n'y avait pas, comme aujourd'hui, un mouvement conjugué de la pratique et la théorie des dictionnaires, cf. Mihindou (2001).

Les données présentées dans les pré-textes du dictionnaire sont des données importantes pour l'acquisition de l'histoire d'une langue. Prenons le cas des pré-textes du *Dictionnaire fang-français et français-fang* produit par Galley (1964), qui présente un texte composé de données sur l'histoire du peuple fang, l'histoire de leur langue et les idées qui ont guidé les recherches précédentes. Ces données sont des données d'arrière-plan (Ekwa Ebanéga 2007: 171). Les étudiants et les élèves pourront tirer là d'importantes informations concernant leur origine et leur passé. L'enseignant expliquera aux élèves et aux étudiants qu'en dehors de la nomenclature (la deuxième composante du dictionnaire) un

traitement additionnel des données du dictionnaire est aussi offert dans un autre texte du dictionnaire. Ils pourront alors accéder à ces données de plusieurs façons.

9.3 Les groupes cibles

La typologie des dictionnaires doit reposer sur l'identification des besoins des utilisateurs cibles, cf. Gouws (2001: 111), Householder (1962: 279). Sur ce point il est important de préciser que le choix d'un dictionnaire doit préalablement faire l'objet d'une analyse détaillée des besoins des utilisateurs ou groupes cibles. La tâche de l'enseignant est, en premier lieu, d'identifier les différents usagers (étudiants ou écoliers) afin de pouvoir arriver à déterminer les besoins qui leur sont propres. Le groupe d'étudiants spécialistes nécessite des dictionnaires spécialisés en rapport avec leur champ d'étude. Dans le domaine de la science, par exemple, il y a des dictionnaires de médecine, de biologie, de physique; dans le domaine des humanités, il y a des dictionnaires de philosophie, d'anthropologie, de sociologie, de littérature. Pour ce qui est de la structuration linguistique du Gabon en général et du fang en particulier, Nzang-Bié (2004) affirme qu'il est nécessaire d'identifier au préalable l'utilisateur cible si l'on veut déterminer la fonction que l'on veut assigner au dictionnaire, car de lui dépend tout le processus du dictionnaire. Dans le cadre du fang, l'auteur distingue trois catégories d'utilisateurs cibles:

- catégorie 1: l'utilisateur cible ne sachant ni lire, ni écrire encore moins parler sa langue;
- catégorie 2: l'utilisateur cible ne sachant ni lire, ni écrire, mais sachant parler sa langue; et
- catégorie 3: l'utilisateur cible sachant lire, écrire et parler sa langue.

Il est à noter que les catégories 1 et 2 identifiées par Nzang-Bié (2004) sont uniquement pour le débutant. Alors que pour la catégorie 3, on peut distinguer le débutant, l'intermédiaire et l'avancé. Le but de l'enseignant est de pouvoir guider les usagers composés d'écoliers, d'étudiants, d'enfants et d'adultes sur le type de dictionnaire à choisir aux seules fins de satisfaire leurs besoins.

Il faut noter qu'à chaque type d'usager correspond un niveau d'apprentissage bien précis.

9.4 Les types de dictionnaires

Dès que l'enseignant connaîtra les besoins spécifiques des usagers, il devra être en mesure de les guider quant au choix des dictionnaires capables de répondre à leurs besoins. Mais avant de discuter du type du dictionnaire proprement dit, il faut savoir que le choix du dictionnaire doit tenir compte du contexte socio-

culturel du pays. Pour le cas du Gabon, l'enseignant orientera les usagers qui veulent apprendre dans la langue étrangère à choisir un dictionnaire bilingue qui comprendrait des informations telles que les équivalences. Il doit guider les usagers qui veulent communiquer dans leur langue à opter pour un dictionnaire monolingue qui apporterait une description assez explicite de la langue. Ce dictionnaire comprendrait des informations telles les définitions, la prononciation, etc. Il résulte de ce qui précède que le type de dictionnaire qui répondrait mieux aux attentes des usagers est le dictionnaire hybride qui englobe les caractéristiques d'un dictionnaire monolingue et celles d'un dictionnaire bilingue. Ce modèle sera intégral et comprendra non seulement le plus grand nombre d'informations linguistiques (définitions, équivalences, etc.), mais il apportera aussi des informations extralinguistiques. Lorsque, par exemple, se posent les problèmes sémantiques entre les élèves de cultures différentes, les illustrations peuvent couvrir le sens du mot.

Brewer (2006) affirme que l'un des problèmes auxquels les lexicographes et la lexicographie font face en ce moment est de savoir comment éduquer les usagers à comprendre les buts de chaque type de dictionnaire et quelles sont les différences entre les types de dictionnaires. Pour ce qui est du Gabon, l'enseignant doit connaître au préalable le but véritable de tout type de dictionnaire et les différences entre les types de dictionnaires. Cette connaissance est nécessaire car elle va permettre à l'enseignant non seulement de guider les étudiants à utiliser et à consulter les dictionnaires convenablement, mais aussi pour conseiller aux écoles comment acheter les dictionnaires qui sont capables de répondre à leurs besoins. La connaissance de la typologie du dictionnaire aidera à élaborer la structure du dictionnaire (Hadebe 2004).

Il a été dit plus haut que le but du dictionnaire est de répondre à un besoin de communication. Pour le cas du Gabon, les enseignants orienteront les étudiants à consulter un dictionnaire qui satisfera leurs besoins. C'est ainsi que lorsqu'ils se trouveront, par exemple, en situation d'apprentissage d'une langue étrangère et seront confrontés par des problèmes de communication au niveau de la production, les enseignants pourront guider les étudiants et les élèves à utiliser un dictionnaire monolingue. Lorsqu'ils seront confrontés par le problème de communication au niveau de la réception, les enseignants les aideront à utiliser un dictionnaire bilingue.

9.5 La structure du dictionnaire

Les enseignants apprendront aux étudiants les différentes structures du dictionnaire (la macrostructure, la microstructure, la structure d'accès, la structure d'adressage etc.) et l'importance de chaque structure du dictionnaire. Les étudiants doivent être en mesure de savoir comment utiliser les textes externes du dictionnaire (les pré-textes et posttextes), de pouvoir interpréter les symboles des abréviations, et des explications qui se trouvent dans ces textes externes du dictionnaire. L'apprentissage des structures du dictionnaire est nécessaire pour

les étudiants car tous les textes du dictionnaire ne sont pas intégrés dans la nomenclature. Les enseignants inculqueront aux étudiants ces textes qui ne sont pas intégrés au dictionnaire pour compléter leur information, cf. Gouws (2001: 123).

9.6 La morphologie

Puisque les dictionnaires, comprennent la formation des mots et leur morphologie, l'enseignant prendra soin d'insister sur les différents éléments qui composent le lemme. L'identification de ces éléments crée un problème important sur la façon de déterminer ce qui devrait être identifié comme lemme et ce qui ne devrait pas. À cet effet, Bwenge (1989: 5) notait:

Le problème central est en particulier la méthode d'arranger les articles nominaux et verbaux de la langue, émanant de la structure morphologique complexe commune aux langues bantoues, d'un système de classification morphologique classant des noms par catégorie au moyen des affixes dérivationnels.

Compte tenu de ce qui précède, Ekwa Ebanéga (2007) a signalé que la présentation des données morphologiques (indication de ce qui sont le préfixe, la tige et le nombre de classe) pour chaque signe de lemme peut être compliquée pour les élèves débutants. Par ailleurs, la décision concernant ce qu'il faut inclure ou exclure concernant des données sur la partie du discours doit être faite selon les besoins et les qualifications de référence des utilisateurs.

Dans les posttextes du *Dictionnaire fang–français et français–fang*, l'auteur présente un texte nommé la minigrammaire dans laquelle les élèves et étudiants pourraient se familiariser avec certains aspects de la structure de la langue fang, à savoir les classificateurs, les préfixes, etc. L'enseignant fournirait aux élèves et aux étudiants des conseils pratiques et des informations sur l'emploi des éléments morphologiques.

9.7 La phonétique

La phonétique est une donnée nécessaire du dictionnaire. Le dictionnaire est non seulement un répertoire de mots, mais également un registre des sons d'une langue. Pour connaître une langue, nous devons apprendre les sons de cette langue. En 1954, Malmberg (cité par Berri et Pagel 2005: 1) soulignait déjà l'importance de la phonétique dans l'apprentissage des langues:

L'enseignement des langues étrangères est aussi un domaine où la phonétique a une très grande importance pratique. Celui qui veut apprendre à bien prononcer une langue étrangère, devra acquérir d'abord la maîtrise d'un grand nombre d'habitudes articulatoires nouvelles (une base articulatoire). Il ne faut pas croire qu'il s'agisse seulement d'apprendre quelques sons nouveaux et, pour le reste, utiliser les sons déjà connus. C'est un système d'habitudes articulatoires, y com-

pris l'intonation et l'emploi des accents expiratoires, qui sera remplacé par quelque chose de nouveau.

Les enseignants montreront aux étudiants les différents alphabets existant en langues gabonaises comme l'Alphabet des Langues du Gabon (ASG) et l'Orthographe des Langues Gabonaises (OLG). La connaissance des alphabets existant en langues gabonaises permettra aux étudiants de connaître les phonèmes qui existent dans les langues gabonaises. Pour ce qui est de la langue fang (ntumu), les étudiants sauront qu'on distingue neuf voyelles (i, e, ε, y, ə, a, u, o, ɔ) et plusieurs consonnes (p, b, m, n, t, d, k, g, f, v, s, z, ʃ, l, r, etc.). La connaissance des variantes allophoniques de certains phonèmes de la langue fang est aussi nécessaire. Par exemple, la consonne /b/ est réalisée [p] en fin de mot ou de phrase et /b/ ailleurs. Exemple:

/ekob/ → [ékəp] "la peau"

/bəfam/ → [bəfam] "les hommes"

Les étudiants devraient aussi se familiariser avec les types de tons qui existent en langues gabonaises. La connaissance de la typologie des tons aidera les étudiants à bien prononcer les mots de la langue. Dans la majorité des langues bantoues en général et les langues gabonaises en particulier, les tons jouent un rôle distinctif dans la mesure où ils permettent de distinguer deux mots. Exemples:

pové: [tóláká] "abandonner (village), émigrer"
[tòláká] "créer, deviner, inventer"

yipunu: [ùtójà] "se fatiguer"
[ùtòjà] "se battre"

fang (atsi): [àbì] "cuisse d'homme"
[àbí] "excrement"

Dans les posttextes, c'est-à-dire, la troisième et la dernière partie du *Dictionnaire fang-français et français-fang*, Galley (1964) passe en revue les signes et les symboles retenus pour transcrire les voyelles et les consonnes du fang. Les enseignants encourageront les élèves et les étudiants à se familiariser davantage avec ces posttextes contenant des signes et des symboles en vue de connaître avec précision les sons des langues étudiées et décrites.

9.8 L'orthographe

L'orthographe dont le but est de décrire le dialecte standard doit prendre en compte l'orthographe standard de l'item lexical dans lequel on trouve des variétés de forme de la langue, cf. Mdee (1990). L'enseignant connaîtra la forme

standard et les règles orthographiques qui sont utilisées dans le dictionnaire. Il insistera sur le fait que l'orthographe revêt une importance dans tous les écrits. Et le but du dictionnaire c'est de permettre aux étudiants de vérifier l'orthographe du mot utilisé.

Pour ce qui est des règles orthographiques des langues gabonaises, nous notons, par exemple, l'harmonisation, l'assimilation des voyelles *u* ou *i* qui deviennent respectivement *w* et *y*. Pour ce qui est du yipunu, par exemple, la voyelle *u* placée devant *a*, *e*, et *i* devient *w*; la voyelle *i* devient *y* lorsqu'elle précède les voyelles *a*, *e*, *o* et *u*. Exemple:

muana → mwana "enfant"

uliomisa → ulyomisa "rendre propre"

mueni → mweni "étranger"

Malgré son importance et sa valeur, l'orthographe est l'une des données les plus mal utilisées. La tâche des enseignants serait très complexe, en ce qui concerne spécialement les données orthographiques; ils devraient en être conscients. Les enseignants conseilleront aux étudiants de se référer aux pré-textes aux seules fins d'avoir des explications relatives aux données orthographiques du dictionnaire.

9.9 La sémantique

Dans l'évaluation critique des dictionnaires existants en langues gabonaises, le sens ne renseigne pas suffisamment sur les différents emplois du mot en contexte. Par conséquent, les usagers natifs auront des problèmes pour comprendre le contexte sémantique dans lequel les mots sont utilisés. À cet effet, le jugement de Tarp (2004: 313) est valide:

Des données sur des significations sont exclusivement fournies en vue de confirmer les usagers quel est le vrai sens du mot. Des usagers sont déjà censés avoir une idée de ce qu'ils veulent exprimer, c'est-à-dire ils connaissent déjà la signification et la nécessité juste des mots. Comme tel il est évident que le dictionnaire devrait exposer toutes les différentes significations et les sens du mot afin de satisfaire aux besoins de l'utilisateur dans n'importe quelle situation.

Puisque la complexité sémantique met en jeu plusieurs niveaux de données, le but du dictionnaire consiste à pourvoir aux élèves et aux étudiants tous les aspects sémantiques des mots, allant du sens des mots, à celui des phrases, aux relations sémantiques entre phrases dans le discours, et aux relations pragmatiques qui mettent en jeu l'utilisation des mots dans les contextes divers. L'enseignant aidera les élèves et les étudiants à vérifier si tous les sens du mot sont utilisés dans le dictionnaire. Cela nécessite au préalable qu'ils soient aptes à utiliser les procédés dictionnaires tels que les renvois qui permettent d'éclair-

rer le sens d'un mot et de connaître les relations sémantiques que le mot a avec les autres mots du dictionnaire. Une fois de plus, l'enseignant est appelé à encourager les élèves et les étudiants à se familiariser avec des renvois (les signes, les abréviations et les symboles) expliqués dans les pré-textes du dictionnaire.

9.10 Le vocabulaire

L'acquisition du vocabulaire est important pour l'apprentissage d'une langue. Les enseignants doivent aider les étudiants et les élèves à acquérir le vocabulaire par le dictionnaire. Ils les guideront dans la connaissance des propriétés ou des caractéristiques du mot. À ce propos Laufer (1992: 71) affirme que "connaître un mot implique connaître toutes ses propriétés (...). Quand une personne connaît un mot, elle connaît la prononciation, l'orthographe, les composantes morphologiques, les comportements syntaxiques dans un énoncé, les situations appropriées pour l'utiliser". Comme le dictionnaire est un livre de mots, il est l'un des moyens le plus efficace d'apprendre le vocabulaire; il permet de prendre le mot directement dans son contexte et il permet aussi aux étudiants et aux élèves de développer une connaissance générale de mots et de groupes de mots dans le but de comprendre les textes littéraires. Lorsque nous voulons apprendre, par exemple, un nouveau mot, nous consultons un dictionnaire pour connaître le mot dans son contexte. Le dictionnaire donne des exemples de phrase pour chaque mot. Pour le parler fang, les étudiants et les élèves pourraient trouver dans le *Dictionnaire fang-français de français-fang* de Galley (1964) plusieurs exemples de phrases accompagnant le mot. Ils pourraient avoir alors plusieurs exemples de contextes d'utilisation.

9.11 L'étymologie

Il a été question plus haut de préciser que la consultation du dictionnaire est une aide pour l'apprentissage du vocabulaire dans l'acquisition d'une langue, car le dictionnaire permet de connaître le mot dans son contexte. À l'instar du vocabulaire, les enseignants devraient aussi savoir attirer l'attention des étudiants et des élèves à connaître l'étymologie des mots. Puisque le dictionnaire donne la possibilité de connaître le mot: "Séparer une langue de son passé, c'est manquer la possibilité de montrer la langue dans son contexte" (Landau 2001: 132). L'étymologie permet donc d'enrichir son vocabulaire et mieux comprendre le sens des mots.

10. Conclusion

Dans ce travail, il a été question de voir dans quelle mesure le dictionnaire peut participer à l'enseignement des langues gabonaises. L'esprit qui a animé cette

contribution était de présenter le contexte multilingue du Gabon, de montrer que les problèmes lexicographiques sont liés à l'éducation au Gabon: les dictionnaires en langues gabonaises sont recherchés, rarement utilisés et trouvés; le gouvernement gabonais ne subventionne même pas l'acquisition des dictionnaires existants pour stimuler les études appropriées; certains usagers ne sont pas informés de l'existence de ces outils.

Pour ce qui est du dictionnaire des langues nationales et selon le vœu exprimé par la majorité des informateurs, un outil d'enseignement comme le dictionnaire doit aller de pair avec l'intégration des langues gabonaises. Leur intégration dans le système éducatif gabonais doit se faire à partir du primaire voire dès la maternelle.

En ce qui concerne la valorisation du dictionnaire dans l'enseignement des langues gabonaises, il faut retenir que l'organisation des conférences et séminaires permettraient au public de prendre conscience de l'existence et de l'importance de cet outil dans le système éducatif.

La mise en place au Gabon des structures de recherche comme les unités lexicographiques serait un facteur déterminant pour la promotion et la sauvegarde des langues gabonaises. Ces unités lexicographiques serviraient d'interface pour la collaboration entre les lexicographes, linguistes, enseignants et apprenants (étudiants et élèves). En établissant les unités lexicographiques au Gabon, ceci aurait comme résultat la production des dictionnaires et la préservation de la diversité linguistique sous toutes ces formes.

L'utilisation de l'outil qu'est le dictionnaire serait effective chez les étudiants et les élèves en visant d'abord la formation des enseignants. Ce dernier transmettra son savoir et son expérience dictionnaire aux apprenants. Une fois que les connaissances linguistiques sont acquises chez les enseignants, on pourrait également compléter leur formation en leur enseignant plusieurs aspects de la lexicographie tels que la métalexigraphie, l'histoire de la lexicographie, la structure du dictionnaire, les types de dictionnaires, groupes cibles, morphologie, phonétique, orthographe, sémantique, vocabulaire, etc.

Notes

1. Le slogan "Donne moi ton dictionnaire, j'évaluerai ta langue" (français) ou "Give me your dictionary, I will evaluate your language" (anglais) né des nombreuses discussions que j'ai eues avec le professeur Gouws lors de la 12^{ème} conférence d'AFRILEX.
2. Le PanSALB est un projet dont le rôle est d'établir les unités lexicographiques pour les langues officielles de l'Afrique du Sud, de favoriser le multilinguisme et développer les langues qui étaient autrefois marginalisées. Grâce à ce projet l'unité lexicographique pour chacune des onze langues officielles de l'Afrique du Sud, à savoir: zoulou, xhosa, afrikaans, sepedi, anglais, tswana, sotho, tsonga, swati, venda et ndebele, a vu le jour.
3. Dans sa recherche, Joubier (2008) discute "les rapports entre la langue maternelle et la langue étrangère dans l'enseignement précoce" et affirme que *dans le cas de la langue qualifiée de maternelle, l'enjeu de son acquisition est vital pour l'enfant qui grandit, se construit psychologiquement et*

se construit une identité en même temps qu'il entre dans la parole et acquiert la maîtrise du langage dans ses relations sociales au contact des autres. La langue dite étrangère est par nature secondaire, sur un plan temporel d'abord, et surtout parce qu'elle apparaît comme dénuée de sens pour l'enfant, car non rattachée dans son environnement de vie à un immédiat communicatif et affectif.

4. L'inventaire linguistique géographico-administratif proposé par Kwenzi-Mikala (1998: 217) contient 62 parlers répartis en 10 groupes:
 - le groupe mazuna: fan-atsi, fan-make, fang-mvaï, fan-ntumu, fan-nzaman et fan-okak;
 - le groupe myene: enenga, ghalwa, mpongwe, nkomi, orungu et okoa;
 - le groupe mekana-menaa: akele, ungom, lisighu, mbanwe, metombolo, seki, tumbidi, shake, wumpfu et lendambomo;
 - le groupe mekona-mangote: ikota, benga, shamay, mahongwe, ndasha et bakola;
 - le groupe membe (ou okande-tsogho): ghetsogo, ghepinzi, kande, ghevovhe, ghehimbaka, ghevhiya, ebongwe et koto-a-kota;
 - le groupe merye: ghisira, ghivharama, ghivhungu, yipunu, yilumbu, yisangu, ngubi, civili, yirimba et yighama;
 - le groupe metye: yinzebi, yitsengi, yimvhele, yivhili, liduma, liwanzi et yibongo;
 - le groupe membre: lembaama, lekanini, lindumu, lateghe et latsitseghe;
 - le groupe makaana: bekwil, shiwa (ou makina) et mwese;
 - le groupe baka
5. Pour l'identification des voyelles et des consonnes en langue fang (ntumu), voir. Ondo-Mébiame (1992).

Références bibliographiques

- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leyde: E.J. Brill.
- Berri, A. et D. Pagel.** 2005. *La phonétique dans la classe de FLE*. <http://www.fdlm.org/fle/article/339/phonetique.php>.
- Bokoko, E.B.** 2004. *Enseignement des langues nationales au Gabon: bilan et perspectives*. <http://www.bibliotheque.refer.org/livre244/124423>.
- Brewer, C.** 2006. *Examining the OED* [online]. <http://www.english.ox.ac.uk/oed>.
- Bwenge, C.** 1989. *Lexicographical Treatment of Affixal Morphology: A Case Study of Four Swahili Dictionaries*. James, G. (Éd.). 1989. *Lexicographers and their Works*: 5-17. Exeter: Exeter University Press.
- Dolezal, F.T. et D.R. McCreary.** 1999. *Pedagogical Lexicography Today: A Critical Bibliography on Learners' Dictionaries with Special Emphasis on Language Learners and Dictionary Users*. Lexicographica. Series Maior 96. Tübingen: Max Niemeyer.
- Ekwa Ebanéga, G.-M.** 2007. *A Microstructural Programme for Dictionaries in Fang*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Ekwa Ebanéga, G.-M. et F. Tomba Moussavou.** 2006. Des considérations sur la lexicographie et la linguistique au Gabon. *Lexikos* 16: 240-250.
- Ekwa Ebanéga, G.-M. et F. Tomba Moussavou.** 2008. A Survey of Dictionary Use of Gabonese Students at Two South African Universities. *Lexikos* 18: 349-365.
- Emejulu, J.D.** 2001. *Lexicographie multilingue et multisectorielle au Gabon: planification, stratégie et enjeux*. Emejulu, J.D. (Éd.). 2001: 38-57.

- Emejulu, J.D. (Éd.).** 2001. *Éléments de lexicographie gabonaise. Tome I.* New York: Jimacs-Hillman.
- Emejulu, J.D et Y. Nzang-Bié.** 1999. *Linguistic Perspectives in Gabon.* Communication présentée à un colloque organisé par le Summer Institute of Linguistics, Université de Dakota du Nord, Grand Forks, États-Unis, 20 juillet 1999.
- Galley, P.S.** 1964. *Dictionnaire fang–français et français–fang, suivi d’une grammaire fang.* Neuchâtel: Henri Messeiller.
- Gates, E.** 1997. A Survey of the Teaching of Lexicography: 1979–1995. *Dictionaries* 18: 66-93.
- Gouws, R.H.** 2001. Formation lexicographique: approches et thèmes. Emejulu, J.D. (Éd.). 2001: 95-134.
- Hadebe, S.** 2004. Improving Dictionary Skills in Ndebele. *Lexikos* 14: 89-104.
- Hausmann, F.J.** 1986. The Training and Professional Development of Lexicographers in Germany. Ilson, R.E. (Éd.). 1986. *Lexicography. An Emerging International Profession:* 101-110. Manchester: Manchester University Press.
- Householder, F.W.** 1962. Summary Report. Householder, F.W. et S. Saporta (Éds.). 1962. *Problems in Lexicography:* 279-282. Bloomington: Indiana University/Le Havre: Mouton.
- Joubier, A.** 2008. *Les rapports entre langue maternelle et langue étrangère.* <http://www.edufle.net/Les-rapports-entre-la-langue>.
- Kipfer, B.A.** 1987. Dictionaries and the Intermediate Student: Communicative Needs and the Development of User Reference Skills. Cowie, A.P. (Éd.). 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985:* 44-54. *Lexicographica. Series Maior* 17. Tübingen: Max Niemeyer.
- Kwenzi-Mikala, J.T.** 1998. Parlers du Gabon: Classification du 11.12.97. Raponda-Walker, A. (Éd.). 1998. *Les langues du Gabon:* 217. Classiques africains. Libreville: Éditions Raponda-Walker.
- Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography.* New York: Cambridge University Press.
- Laufer, B.** 1992. Corpus-Based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, H. et al. (Éds.). 1992. *EURALEX '92 Proceedings I–II: Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland:* 71-76. Tampere: Université de Tampere.
- Mabika Mbokou, L.** 2001. Le rôle du dictionnaire dans le système éducatif. Emejulu, J.D. (Éd.). 2001: 206-222.
- Malmberg, B.** 1954. *La phonétique.* Paris: Presses Universitaires de France.
- McArthur, T.** 1986. *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablets to the Computer.* Cambridge: Cambridge University Press.
- McKean, E.** 2000. Dictionary Activities in the Elementary Classroom: News for Lexicographers. *Dictionaries* 21: 81-89.
- Mdee, J.S.** 1990. Dictionaries and the Standardization of Spelling in Swahili. *Lexikos* 9: 119-134.
- Mihindou, G.-R.** 2001. Apports des missionnaires à la lexicographie gabonaise: Dictionnaires bilingues fang–français/français–fang; français–yipounou/yipounou–français; français–mpongwé. Emejulu, J.D. (Éd.). 2001: 7-37.
- Moyo, T.** 2002. Mother Tongues versus an Ex-colonial Language as Media of Instruction and the Promotion of Multilingualism: The South African Experience. *South African Journal of African Languages* 22(2): 149-160.

- Ndinga-Koumba-Binza, H.S.** 2005. Politique linguistique et éducation au Gabon: un état des lieux. *Journal of Education* 4(1): 65-78. Réduit: Mauritius Institute of Education.
- Nyangone Assam, B.** 2001. La lexicographie dans l'apprentissage des langues gabonaises. Emejulu, J.D. (Éd.). 2001: 187-205.
- Nyangone Assam, B. et P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Nzang-Bié, Y.** 2004. *La problématique de la révision des dictionnaires anciens: le cas du Dictionnaire fang-français et français-fang de Samuel Galley (1964)*. Communication non publiée. Libreville: GRELACO, Université Omar Bongo.
- Ondo-Mébiame, P.** 1992. *De la phonologie à la morphologie du fang-ntumu parlé à Aboumézok (bantu A.78)*. 2 volumes. Thèse de doctorat (NR). Bruxelles: Université Libre de Bruxelles.
- Tarp, S.** 2004. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.
- Tomba Moussavou, F.** 2007. *Metalexigraphic Criteria for a Monolingual Descriptive Dictionary Presenting the Standard Variety of Yipunu*. Thèse de doctorat non publiée. Stellenbosch: Université de Stellenbosch.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K (Éd.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

Does Johnson's Prescriptive Approach Still Have a Role to Play in Modern-Day Dictionaries?*

Rufus H. Gouws (*rhg@sun.ac.za*)
and

Liezl Potgieter (*liezlp@vodamail.co.za*), *Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa*

Abstract: Samuel Johnson's dictionary (1755) confirmed both the status of dictionaries as authoritative sources of (linguistic) knowledge and the prescriptive approach in lexicography. This approach prevailed for a long time. During the last decades the descriptive approach came to the fore, aptly supported by the increased reliance on lexicographic corpora. Modern-day lexicography has also witnessed the introduction of a third approach, i.e. the proscriptive approach, which includes features of both the prescriptive and the descriptive approach. This article investigates the occurrence of the prescriptive, descriptive and proscriptive approaches in modern-day dictionaries. A distinction is made between dictionaries focusing on language for general purposes and dictionaries focusing on languages for special purposes. It is shown that users rely on dictionaries as prescriptive reference sources and expect lexicographers to provide them with an answer to the specific question that prompted the dictionary consultation process. It is argued that knowledgeable dictionary users must be able to achieve an unambiguous retrieval of information and must be able to rely on the dictionary to satisfy their specific cognitive or communicative needs. Here the prescriptive approach plays an important role.

Keywords: COGNITIVE FUNCTION, COMMUNICATION FUNCTION, CULTURE-DEPENDENT, DESCRIPTIVE, EXACT PROSCRIPTION, EXCLUSIVE PROSCRIPTION, LSP DICTIONARIES, LSP DICTIONARIES, NON-RECOMMENDED FORM, PRESCRIPTIVE, PROSCRIPTIVE, RECOMMENDATION, TYPES OF USERS, USER PERSPECTIVE.

Opsomming: Het Johnson se preskriptiewe benadering nog 'n rol te speel in moderne woordeboeke? Samuel Johnson se woordeboek (1755) het die status van woordeboeke as gesaghebbende houers van (taalkundige) kennis, maar eweneens die preskriptiewe benadering in leksikografie gevestig. Hierdie benadering het lank gegeld. Gedurende die onlangse dekades het die deskriptiewe benadering op die voorgrond getree, sterk ondersteun deur toenemende benutting van korpora. Moderne leksikografie het 'n derde benadering beleef, te wete die proskriptiewe benadering wat kenmerke van sowel die preskriptiewe as die deskriptiewe benadering bevat. Hierdie artikel ondersoek die voorkoms van die preskriptiewe, deskriptiewe en pro-

* This article is based on a paper read at the XVII Biennial Meeting of the Dictionary Society of North America, 26–30 May 2009, Indiana University, Bloomington, United States of America in a session to commemorate the tercentenary of Samuel Johnson's birth.

skriptiewe benadering in moderne woordeboeke. 'n Onderskeid word gemaak tussen algemene en vakwoordeboeke. Daar word aangetoon dat gebruikers op woordeboeke staatmaak as preskriptiewe naslaanbronne en dit van leksikograwe verwag om aan hulle antwoorde te verskaf op die spesifieke vrae wat tot die woordeboekraadpleging aanleiding gee. Kundige woordeboekgebruikers moet daartoe in staat wees om 'n ondubbelsinnige ontsluiting van inligting te bereik en hulle woordeboek te kan vertrou vir die bevrediging van spesifieke kognitiewe en kommunikatiewe behoeftes. Hier speel die proskriptiewe benadering 'n wesenlike rol.

Slutelwoorde: AANBEVELING, ALGEMENE WOORDEBOEKE, DESKRIPTIEF, EKSKLUSIEWE PROSKRIPSIE, GEBRUIKERSPERSPEKTIEF, GEBRUIKERSTIPES, KOGNITIEWE FUNKSIE, KOMMUNIKATIEWE FUNKSIE, KULTUURGEBONDE, NIE-AANBEVOLE VORM, PRESIESE PROSKRIPSIE, PRESKRIPTIEF, PROSKRIPTIEF, VAKWOORDEBOEKE

When I took the first survey of my undertaking, I found our speech copious without order, and energetick without rules: wherever I turned my view, there was perplexity to be disentangled, and confusion to be regulated; choice was to be made out of boundless variety, without any established principle of selection; adulterations were to be detected, without a settled test of purity; and modes of expression to be rejected or received, without the suffrages of any writers of classical reputation or acknowledged authority.

Samuel Johnson: Preface to *A Dictionary of the English Language*

1. Introduction

Whenever reference is made to Samuel Johnson's dictionary one of the most frequently referred to definitions is his description of a lexicographer as a harmless drudge. As a lexicographer in his own right Samuel Johnson imposed the status of a dictionary as an authoritative source of, especially, linguistic information. Looking at the ideals he set for himself in his well-known *The Plan of a Dictionary of the English Language* (1747) his passion for "fixing the language", the so-called "linguistic authoritarianism" puts the lexicographer in a dominant position with the potential to do both good and harm to the language. This takes the lexicographer well beyond the scope of harmless drudgery. We know that Johnson eventually realised that he could not actually achieve the fixing of the language to the extent especially other authors, e.g. Swift, originally had envisaged. Yet, the criteria according to which he selected words for inclusion as lemmata in his dictionary, the way in which he treated some of the words and the way in which he omitted certain words from his dictionary give a clear indication of a certain prescriptive approach — an approach that elevates the lexicographer to the level of being a custodian but also a judge of the language. Where such a prescriptive approach succeeds in assisting the users of a given dictionary in their quest for correct language use one can hardly object to it. However, where such an approach distorts the language by resulting in a subjective and biased presentation of data, a dictionary

becomes a dangerous instrument. In *The Plan of a Dictionary of the English Language* (1747: 4) Johnson says: "It was not easy to determine by what rule of distinction the words of this dictionary were to be chosen. The chief intent of it is to preserve the purity and ascertain the meaning of our English idiom." Where the nature and extent of such a prescriptive approach impedes a dictionary to reflect the actual language use and where it can be seen as an attempt to purify the language in terms of the lexicographer's world-view, more care should be taken before adhering to such an approach.

In spite of all the legitimate criticism of a prescriptive approach and in spite of alternative approaches that might be more feasible, lexicographers should always realise that the typical dictionary user regards a dictionary, *the dictionary*, as an authoritative source of data from which they want to retrieve the information needed to solve those specific problems that prompted the consultation process. In some instances users may desire to rely on a dictionary that is the product of a Johnsonian approach, as stated in his *Plan* (Johnson 1747: 11): "one great end of this undertaking is to fix the English language." Acknowledging that this may still be the case in some instances today, this article looks at different aspects of prescriptive, descriptive and proscriptive approaches in lexicography in an attempt to find a way and to make suggestions that would meet the needs of the intended target users of a given dictionary and satisfy the genuine purpose of that dictionary. What remains non-negotiable is the need for an approach that is scientifically based. In this regard Johnson's remark "in lexicography, as in other arts, naked science is too delicate for the purposes of life" (Johnson 1747: 4) cannot be accepted.

2. Moving from prescriptive to descriptive

The era following the Renaissance was characterised by dictionaries trying to open new worlds to their potential users, with the target users no longer an elite group of scholars or academics as typically seen during the medieval era but rather the average person on the street. In the words of Robert Cawdrey in his *A Table Alphabeticall* (1604):

A Table Alphabeticall, conteyning and teaching the true writing, and understanding of hard usual English wordes, borrowed from the Hebrew, Greeke, Latine, or French. &c.

With the interpretation thereof by *plaine English words, gathered for the benefit & helpe of Ladies, Gentlewomen, or any other unskilfull persons.*

In 1615 William Bathe published his *Ianua Linguarum* (The Gate of Tongues) and this was followed in 1631 by Amos Komensky who went one step further with his *Ianua Linguarum Reserata* (The Gate of Tongues Unlocked). Dictionaries had to unlock the gate of tongues so that unskilful persons could gain access to knowledge. This approach helped to establish the idea of dictionaries being regarded as authoritative sources of linguistic information and Johnson's vision

of employing his dictionary to fix the language firmly ascertained this prescriptive assignment of dictionaries. Agreement with this approach was echoed by the Lord Chesterfield, when commenting on Johnson's dictionary, saying: "Toleration, adoption and naturalization have run their lengths. Good order and authority are now necessary." Over the years different forms of prescription have developed (cf. Bergenholtz 2003) but they are usually all characterised by the lexicographer imposing his/her point of view on the user.

Many lexicographers became more Johnsonian in their prescriptive approach than Johnson himself. This often resulted in an approach characterised by efforts to purify the language, especially in multilingual environments where language contact and the subsequent language influence is an everyday reality. Such a purist approach played a major role in, for example, the early development of Afrikaans when lexicographers used their dictionaries as instruments of lexical ethnic cleansing, trying in vain to rid Afrikaans from the influence of English or, in one of the first Afrikaans dictionaries, i.e. Changuijn's *Proeve van Kaapsch Taaleigen* (1844) where the author endeavoured to rid the then standard Dutch from the emerging Afrikaans. An important paradigm shift was introduced by Philip Gove in *Webster's Third New International Dictionary* (1961) when he clearly stated that a dictionary should reflect language, not set its style. The descriptive approach was later firmly established during the corpus era with lexicographers relying on real language data to produce dictionaries, like John Sinclair's *COBUILD* that clearly states its aim as "HELPING LEARNERS WITH REAL ENGLISH".

As is the case with prescription, different types of description can also be distinguished (cf. Bergenholtz 2003). The nature of a descriptive approach determines that a dictionary should not enforce a lexicographer's biased opinion about what belongs in a dictionary or how a specific word should be treated on the users of that dictionary. A dictionary needs to mirror actual language usage and the full spectrum of lexical items should be regarded as potential candidates for a lemma list. The typological classification of a dictionary will determine the nature and extent of items to be included as macrostructural elements as well as their microstructural treatment. Not only the standard variety of a language but all varieties should be considered for lexicographic exposure and where different variants co-occur in the language it needs to be reflected in a dictionary. In this regard the use of a balanced and representative corpus plays an important role in assisting lexicographers to identify their lemma candidates and also the different senses, uses, morphological features, orthographic variants, etc. of a given word in order to ensure a successful descriptive presentation (cf. Tarp and Gouws 2008). A corpus will also indicate the relevant usage frequency and its implications for the lexicographic process.

Having to decide on a prescriptive or a descriptive approach any lexicographer needs to realise that dictionaries are compiled as utility instruments for specific target users with specific needs and reference skills, consulting their

dictionaries in specific situations of usage. This has to determine each and every aspect of each and every dictionary. Prior to the compilation of a dictionary the lexicographer needs to know what the genuine purpose and the lexicographic functions of the envisaged dictionary will be. The various structures of the dictionary should then ensure that the user will have access to the required data and will be able to achieve an optimal retrieval of information.

3. The user perspective

Lexicographers typically work with the assumption that they are familiar with their intended target users and the needs of these users. The dictionary user is no longer the well-known unknown. Questions may be asked regarding the validity of many lexicographers' claims that they know what their users need. The notion of user-friendliness is unfortunately too often determined by the perspective of the lexicographer and not the perspective of the user. An early warning in a comparable regard was formulated by Philip Gove (1966: 183) when speaking about self-explanatory compounds and stressing that the *self-* in *self-explanatory* should refer to the interpreter of the word and not the word as such. The decision to include these words in a dictionary should be based on what the user and not the lexicographer will regard as self-explanatory. However, lexicographers usually are spot on when they proclaim that their users want an unambiguous answer to the problem that prompted a specific dictionary consultation. Haas (1967: 48) already maintained: "A good dictionary is one in which you can find the information you are looking for — preferably in the very first place you look." One can add to this that users not only need to find data but they also need to retrieve the relevant and correct information. Owing to different needs of the respective user groups the paraphrase of meaning given for the same word will be different in a general dictionary, a specialised dictionary for lay persons or learners at school and a specialised dictionary for experts. In each case the definition has to be relevant, complete and correct but in terms of the needs and reference skills of the specific target users (cf. Bergeholtz and Gouws 2007).

Average dictionary users see a dictionary as a reference source in which they can find solutions in terms of communication needs, i.e. the text reception, text production and translation functions of the dictionary, as well as cognitive needs, e.g. finding the etymology of a word, an indication of the distance of a marathon or the date of birth of Madonna. They do not want to be confronted with choices but want a straightforward answer to their questions — and the answer must be valid, appropriate, correct and relevant in terms of the type of dictionary and the specific user needs. Where a descriptive approach results in the dictionary offering an unqualified variety of options users become confused and often question the ability of the specific dictionary to be authoritative. Even when the variants have equal status and official recognition users rather prefer to find a single form that can be regarded as the only correct form. To illustrate

this: the Afrikaans Language Commission, a commission of the South African Academy for Science and Arts officially entrusted with formulating the rules and regulations for Afrikaans orthography, recognised a number of orthographic variants in their official publication the AWS, i.e. the *Afrikaanse woordelys en spelreëls* (Afrikaans word list and spelling rules). These variants include *ver/vêr* (far); *weereens/weer eens* (once again); *opheterdaad/op heter daad* (red-handed). The AWS is a prescriptive publication presenting the forms acknowledged and prescribed by the Language Commission. However, mother-tongue speakers of Afrikaans consulting this publication as well as dictionaries where these variants are presented often complain that they are looking for a single correct form and do not want a choice between more than one variant. In this regard their ideal dictionary would follow a prescriptive approach by nominating only a single form as the only correct one.

4. Towards a proscriptive approach

In the preface to his dictionary, Samuel Johnson said: "every language has ... its improprieties and absurdities, which it is the duty of the lexicographer to correct or proscribe." Since then the word *proscribe* has acquired term status. In modern-day lexicography, the notion of proscription (from the Latin *proscribere* — to make public) is introduced in Bergenholtz (2001) and further discussed in, among others, Bergenholtz (2003) and Tarp and Gouws (2008). The characteristic feature of a proscriptive approach is that it offers a presentation, based on a recommendation by the lexicographer. This is not a form prescribed by the lexicographer but recommended by him/her as the preferred form, based on his/her conviction or choice. In some cases of proscription, only a single form is presented, the only recommended form, and in some cases more than one form is presented but accompanied by a clear recommendation by the lexicographer as to the preferred form. Proscription may even see the recommendation of more than one form as being equal. Employing an approach where a recommendation is made, distinguishes proscription from description where variant forms are given without an indication of the preferred form. Where only one form exists, the proscription will not differ from prescription. However, contrary to the strong version of prescription, proscription, even where only a single form is recommended, does not have language purification as its aim but rather a reflection of actual language use. In a certain sense proscription could be seen as the best of both worlds of prescription and description, but it is actually much more radical. In recommending a specific form, the proscripting lexicographer does not have to adhere to the rules or regulations of a normative or prescribing body. The lexicographer may apply his/her own criteria, e.g. general usage frequency or the choice of informed users, to determine the recommended form. It is important that the specific nature of the proscriptive approach employed in a given dictionary, i.e. the motivation for recom-

mending specific forms, needs to be explained in the users' guidelines text of that dictionary.

Proscription implies a distinction between recommended and non-recommended forms. According to an approach of *exclusive proscription* only the recommended form will be included in a given dictionary. Contrary to this approach, *exact proscription* (cf. Bergenholtz 2003: 77) sees the recommendation of only one form but the other variants may be mentioned in the dictionary. Tarp and Gouws (2008) discuss various aspects regarding the implementation of a proscriptive approach, including different ways in which the non-recommended forms can be presented. This regards both the macro- and micro-structural representation. On a macrostructural level, the non-recommended lexical items should be included as lemmata that are guiding elements of cross-reference articles, referring the user to the article in which the recommended form is given as lemma sign. Such a procedure would ensure that users consulting a dictionary to find a non-recommended form will be guided to the recommended form. However, in terms of the descriptive nature of a proscriptive dictionary and the assignment to reflect actual language use and not only the pure or correct language a dictionary should also inform the user who goes directly to the recommended form of the existence of alternative forms. These forms should also be given in the article of the lemma that represents the recommended form. This can be done in various ways and at present dictionaries actually have two competing ways of presenting variants, with the type of variant determining the article slot where they should be included. Orthographic variants are typically presented in the comment on form, in close vicinity to the lemma sign, cf. the following example from *The New Oxford Dictionary of English* (NOED):

eirenicon /.../ (also **irenicon**) ...

The word *irenicon* has also been included as lemma but with a restricted treatment, indicating to users that this word is "a variant spelling of **EIRENICON**". This well-established convention of presentation already displays a proscriptive approach with the lexicographer implicitly indicating that *eirenicon* is the preferred form by allocating the full treatment to this word. By also including the non-recommended variant as a lemma the descriptive component of the proscriptive approach is ascertained. Lexical-semantic variants, e.g. synonyms, are usually presented in a slot in the comment on semantics. In the comprehensive Afrikaans dictionary, the WAT, i.e. *Woordeboek van die Afrikaanse Taal* (Dictionary of the Afrikaans Language), the lemma *katoensnywurm* (a certain type of cutworm on cotton plants) receives a full treatment, including a paraphrase of meaning. In the last search zone in the comment on semantics of this article, a listing of synonyms is presented: *boor-*, *klim-*, *tabak-*, *snywurm*. From this article the user can deduce that *katoensnywurm* is regarded by the lexicographer as the recommended form but that the other synonyms also prevail as Afrikaans lexical items. By including them the dictionary reflects an aspect of the actual state

of the lexicon of Afrikaans. Where none of the lexical variants of a recommended form are rejected, the lexicographer should make sure that the prescriptive approach sees the inclusion of these forms not only as individual macrostructural items but also as microstructural items in the article of the lemma sign representing the recommended form, as is the case in the above-mentioned presentation in the WAT. The NOED includes the lemma sign *ejector seat* but allocates a limited treatment to its article and cross-refers the user to the recommended variant:

ejector seat ... another term for EJECTION SEAT

The article of the lemma sign *ejection seat* accommodates the full treatment but no indication is given of the variant form *ejector seat*. This is a less successful form of proscription, i.e. *monodirectional proscription*.

Besides the distinction between recommended and non-recommended forms, the application of proscription should also take cognisance of the distinction between accepted and rejected non-recommended forms. Where the lexicographer regards a non-recommended form as one that should actually be rejected by the language users the prescriptive approach allows the inclusion of such a rejected form in the dictionary in order to mark the word as a candidate for rejection or to ascertain its disallowance. The HAT, i.e. *Verklarende Handwoordeboek van die Afrikaanse Taal* (Explanatory Desk Dictionary of the Afrikaans Language) includes the word *komper* and gives the word *rekenaar* (computer) as its synonym but complements this restricted treatment with a note indicating that the word *komper* is not acknowledged by either experts in the field of computers or by the Afrikaans Language Commission. In the article of the lemma sign *rekenaar*, no reference is made to the non-recommended form *komper*. The user confronted with the word *komper* will be able to find the word in the dictionary and obtain the necessary guidance that it is not an acknowledged form along with a reference to the correct or recommended form. The user consulting the dictionary for a retrieval of information regarding the word *rekenaar* does not need to see the non-allowed form and therefore, contrary to the way in which allowed non-recommended forms are treated, no indication is given of the word *komper* in the article of the lemma sign *rekenaar*. Here monodirectional proscription is the acceptable procedure. Including and marking rejected forms is not something new. In his *Plan*, Samuel Johnson (1747: 29) already said: "Barbarous or impure words and expressions, may be branded with some note of infamy, as they are carefully to be eradicated wherever they are found."

Proscription gives the lexicographer the freedom to recommend one or more forms, to acknowledge those forms that are acceptable albeit that they are not recommended by the lexicographer and to make dictionary users aware of disapproved, rejected or even forbidden forms. Compared to a prescriptive or a descriptive approach, proscription allows a much more balanced reflection of the actual language.

The macrostructural application of proscription does not only target main lemmata. Sublemmata also fall within its scope. Fixed expressions represent one type of lexical item, typically included as a secondary treatment unit in dictionaries (cf. Potgieter 2008, Gouws To appear). Fixed expressions often have variants and all these variants are not equally acceptable. As an equivalent of the English idiom *blood is thicker than water* Afrikaans has the expression *waar bloed nie kan loop nie, daar kruip dit* (literally: blood crawls where it cannot walk). This should be given as the recommended variant but a frequent variant is the direct translation from English *bloed is dikker as water*. A proscriptive approach should see the inclusion of this variant but as a non-recommended form.

Proscription does not only have lexical items included as treatment units in its scope. Its application can also be directed at microstructural entries presented in the treatment of a given lexical item, e.g. different pronunciation and morphological forms, different uses of a given word, different collocations, etc. The Afrikaans translation of *at a stage is in 'n stadium*, with the noun *stadium* collocating with the preposition *in*. A frequently used but less correct form is the collocation *op 'n stadium*. In the article of the lemma sign *stadium*, the co-text slot should include the collocation *in 'n stadium*, clearly marked in terms of the domestic convention of the specific dictionary as the recommended form, but, in a proscriptive way, the collocation *op 'n stadium* could also be given as a microstructural entry. A proscriptive approach may merely give the collocation *op 'n stadium* as an alternative, albeit the non-recommended one. A stronger version of proscription may tend to be more prescriptive by explicitly indicating that the non-recommended form should be forbidden. This is in accordance with a point of view that dictionaries should also guide their users by including as marked entries some disapproved forms.

For a dictionary adhering to a proscriptive approach, it is important that the lexicographers should negotiate the best possible article slot for non-recommended forms, presented as microstructural entries in the article of a lemma representing a recommended form. By listing the variants in the comment on form in the article slot immediately following the lemma sign or the items presenting grammatical data, the variants occupy a position of salience that might obstruct the rapid inner access that should lead the user to the comment on semantics. Where both orthographic and lexical variants are given in the same article, it could be confusing to enter one type of variant in the comment on form and the other in the comment on semantics. A more acceptable way may be to introduce one article slot dedicated to non-recommended variants. This could be done by employing an extended obligatory microstructure with such an article slot for non-recommended variants given as a final search zone that follows the default article structure. By means of structural indicators, such a search zone should be clearly marked so that it can be easily identifiable as the search zone accommodating non-recommended variants. These variants can be of the lexical item represented by the lemma sign or of

any other entry presented as microstructural item. Within such a zone for non-recommended forms provision could be made for different subzones to accommodate the different types of variants. By only accessing the obligatory microstructure, the dictionary seems to follow an approach of *exclusive proscription* that could easily be regarded as prescription. By merely looking at the slot in the extended obligatory microstructure, the dictionary seems to follow a descriptive approach. Looking at the entire article, one realises that a proscriptive approach has been employed. This type of presentation can enhance the extent of the information transfer that can be achieved in a given dictionary.

5. Valid applications of a prescriptive approach

Modern-day lexicographic theory is not prescriptive by imposing one theoretical model on all dictionaries. Different typological categories demand different structures, contents and functions. Similarly the application of procedures of prescription, description and proscription should be determined by the specific dictionary type, the needs of the users and the situation of use.

Within the typological classification of dictionaries, the distinction between LGP dictionaries (language for general purposes) and LSP dictionaries (language for special purposes) has implications for various aspects of the theoretical models. As mentioned before, dictionaries are tools made for the purpose of fulfilling specific user needs (Bergenholtz and Nielsen 2006: 283), and LGP and LSP dictionaries make provision for distinctly different needs. Quite often general dictionaries do include a limited number of terms from specialised fields. This is due to various reasons, e.g. the lack of relevant LSP dictionaries or the need of users of LGP dictionaries to have access to those terms. According to McAdam and Milne (1982), at the time when Johnson compiled his dictionary, technical dictionaries were often more comprehensive and accurate than general dictionaries. Consequently Johnson drew heavily on these dictionaries, e.g. for legal, medical and ecclesiastical terms. Today this still applies in languages with an insufficient typological infrastructure. Access is also needed in general dictionaries to those terms typically used in conversation between the expert and the lay person, e.g. medical or legal terms. In general dictionaries, the treatment of technical terms must ensure an unambiguous retrieval of information and this demands that the paraphrase of meaning is formulated for lay persons and not subject field experts.

Even in LSP dictionaries provision needs to be made for different user groups, with the users being divided into three groups: experts, semi-experts and lay people (cf. Bergenholtz and Tarp 1995 and De Foglio and Lubbe 2002). The expert is usually a subject specialist and is familiar with the terminology. The semi-experts are often students of a subject and have a basic knowledge of the subject and the terminology. The last group, the lay people (who also include translators) often have very little or no knowledge of the subject and the terminology. The expert is someone who might be interested in finding

variant forms for a given term. A proscriptive approach with a recommended form supplemented by the inclusion of variants will respond to the needs of this expert user. In the *Fachwörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* currently being compiled the article of the lemma sign *Akronymzuordnungsangabe* (item relating an acronym to its full form) gets a full treatment, including an article slot, preceded by the non-typographical structural indicator "=" in which the lexical variant *Kurzwortzuordnungsangabe* is given — as a non-recommended but permissible variant. Proscription allows a better data transfer than prescription would have achieved. However, whereas the expert might be interested to find variant forms for a given term the lay person typically needs one form that can be used without the fear of making a mistake. These users need active guidance from the lexicographer telling them that a given word is the appropriate one for a specific technical context. A form of prescription, giving unambiguous guidance with relation to the appropriate, accepted and correct form is needed. This can be done by means of prescription but exclusive proscription could have the same result. This is a section of the lexicographic practice where prescription can be a viable option and it is therefore the one field in modern-day lexicography where lexicographers still often make use of a prescriptive approach in bilingual dictionaries by providing only one translation equivalent for a given word/term. The dictionary user therefore has no trouble deciding on a translation equivalent since the lexicographer has already made the decision. If the lexicographer makes use of a less prescriptive approach and provides a number of translation equivalents in the dictionary article, the experts will usually have sufficient knowledge of the subject field to be able to decide which translation equivalent to choose for a specific text or context. Semi-experts and lay people will however most probably need help in the form of glosses, labels or notes in order to decide on the correct equivalent.

LSP dictionaries do sometimes have a form of descriptive assistance by giving both the British and the American English spelling of a term, but not much more. Once again, for the lay user this may suffice. Experts and semi-experts might need more. This is especially true in LSP dictionaries dealing with terms for which culture-dependent variants occur in the specific language. Sepedi, one of the eleven official languages of South Africa and a member of the Sotho language family, has a range of traditional medical and health care terms. In an LSP dictionary dealing with this field, it is important that the traditional variants should be provided along with the official Western terms. Patients often come to the medical doctor and use the traditional term to explain their problem. As an example: when they have abdominal pain, some Sepedi speakers would say that they have been bitten by a snake. Ignorant doctors had treated such patients with snake bite serum which did not solve the problem. An LSP dictionary of medical and health terms should give the recommended term but also indicate the prevailing variant.

The term *prescriptive* carries a lot of baggage. Its application does not only

imply an indication of a single correct form or variant. It also has the connotation of subjectivity, bias and the lexicographer imposing his views and interpretations on the dictionary users. As indicated earlier, proscription can be seen as offering the best features of both prescription and description. Different types of proscription prevail with, for example, exclusive proscription allowing the lexicographer to present only one form without reference to any alternative form. This represents the form recommended by the lexicographer but not on the basis of bias, subjectivity and attempts to purify the language. These issues should in any case not come to the fore in any lexicographic endeavour. Other forms of proscription allow the lexicographer to give a much more comprehensive account of different categories of variants.

6. In conclusion

As a dynamic discipline the development of lexicography is characterised by trends and changes. This applies to both the theoretical and the practical component. As an example of trends in theoretical lexicography, one can refer to the shift in research focus from the linguistic contents of dictionaries to the structure of dictionaries and then to lexicographic functions. In the lexicographic practice, one such trend has been the introduction of a corpus-based approach. Johnson's prescriptive approach can be seen as a trend in the lexicographic practice that formed the basis for different realisations and this evolution created the need for a subsequent trend, i.e. the descriptive approach, followed at a later stage by the proscriptive approach. Just as the focus on lexicographic functions does not eschew the value of the linguistic contents of dictionaries or dictionary structures, the proscriptive approach does not imply a total abolishment of prescriptiveness. Although not aiming at fixing the language, a proscriptive approach makes a recommendation to the user and this is typically based on the most appropriate form for a given situation of usage. In addition, the typological expansion in the lexicographic practice resulted in some dictionaries planned and compiled to fulfil a normative assignment which strongly resembles prescription. This is especially true not only in LSP dictionaries where users often are in need of the correct form but also in, for example, school dictionaries where the users need to find the officially recognised and standardised variants of a given language.

By their nature many types of dictionaries can be regarded as normative and they will therefore always maintain some aspects of a prescriptive approach. In order to convey their data in an unbiased way, dictionaries no longer should present the subjective attempts of a lexicographer to purify the language or to portray a specific ideological, religious or political point of view but rather opt for a form recommended on account of its active occurrence in the real language use. In the majority of articles in his dictionary, Samuel Johnson has done exactly that, employing a form of prescriptivism that is today regarded as exclusive proscription.

Hulbert (1960: xiv) so aptly remarked that "the authority of dictionaries is only that of the men who made them". If that authority is misused to reflect the biased perspective of the dictionary compilers, the prescriptive lexicographer influences language in an unrepresentative way and becomes a harmful drudge. If, however, that authority combines the best of Johnson with Gove's attempts to reflect the real language and the needs of users to find a recommended form representing appropriate and correct language use, the lexicographer gives the intended target users a utility tool that can enhance their communication and cognitive skills.

References

- Bathe, W.** 1615. *Ianua Linguarum*.
- Bergenholtz, H.** 2001. Proskription, oder: So kann man dem Wörterbuchbenutzer bei Textproduktionsschwierigkeiten am ehesten helfen. Lehr, A. et al. (Eds.). 2002. *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik. Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*: 499-520. Berlin/New York: de Gruyter.
- Bergenholtz, H.** 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Bergenholtz, H. and R.H. Gouws.** 2007. Korrek, volledig, relevant. Dit is die vraag aan leksikografiese definisies. *Tydskrif vir Geesteswetenskappe* 47(4): 568-586.
- Bergenholtz, H. and S. Nielsen.** 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Botha, W.F. et al. (Eds.).** 1951-. *Woordeboek van die Afrikaanse Taal*. Stellenbosch: Buro van die WAT.
- Cawdrey, R.** 1604. *A Table Alphabeticall*. London: E. Weaver.
- Changuion, A.N.E.** 1844. Proeve van Kaapsch Taaleigen. Changuion, A.N.E. 1844. *De Nederduitsche Taal in Zuid-Afrika hersteld*. Cape Town: J.H. Collard.
- Chesterfield, Lord.** 1754. Letter to *The World*.
- De Foglio, F. en H.J. Lubbe.** 2002. 'n Praktiese benadering tot die samestelling van 'n tweerigtingvakwoordeboek met Afrikaans en Italiaans as taalpaar. *Lexikos* 12: 117-132.
- Gouws, R.H.** To appear. Fixed Word Combinations as Second Level Treatment Units in Dictionaries. Ďurčo, Peter (Ed.). To appear. *Proceedings: International Colloquium on Lexicography, Bratislava, 2008*.
- Gove, P.B.** 1966. Self-explanatory Words. *American Speech* 41: 182-198.
- Gove, P.B. (Ed.).** 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: Merriam-Webster.
- Haas, M.R.** 1967. What Belongs in a Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 45-50. Bloomington: Indiana University.
- Hulbert, J.R.** 1960. *Dictionaries: British and American*. Second Edition. London: André Deutsch.
- Johnson, S.** 1747. *The Plan of a Dictionary of the English Language*. Facsimile Edition 1970. Menston: The Scholar Press.

- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton/T. & T. Longman et al.
- Komensky, A.** 1631. *Ianua Linguarum Reserata*.
- McAdam, E.L. and G. Milne.** 1982. *Johnson's Dictionary. A Modern Selection*. London: Papermac.
- Odendal, F.F. and R.H. Gouws.** 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Fifth Edition. Cape Town: Pearson Education.
- Pearsall, J. (Ed.)** 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Potgieter, L.** 2008. 'n Moontlike model vir die bewerking van idiome in tweetalige woordeboeke. *Lexikos* 18: 190-206.
- Sinclair, J. (Ed.)** 1988. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- Tarp, S. and R.H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Taalkommissie van die S.A. Akademie vir Wetenskap en Kuns.** 2001. *Afrikaanse Woordelys en Spelreëls*. Cape Town: Pharos.
- Wiegand, H.E. (Ed.)** 2010. *Fachwörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*. Berlin: de Gruyter.

Firming Up the Foundations: Reflections on Verifying the Quota- tions in a Historical Dictionary, with Reference to *A Dictionary of South African English on Historical Principles**

Sheila Hicks, *Dictionary Unit for South African English, Rhodes University,
Grahamstown, South Africa (s.hicks@ru.ac.za)*

Abstract: *A Dictionary of South African English on Historical Principles* (DSAEHist) is rooted in quotation evidence. It contains just over 8 000 South African English entries, with about 45 000 citations to support those words included as headwords in the dictionary. Using the legacy electronic format in which DSAEHist was typeset, the Dictionary Unit for South African English embarked on a digitising process of DSAEHist, during which it became clear that the quotations would benefit from a full review involving the verification of all quotations against their original sources. This article examines the evolution of the quotation verification project from its beginnings as an entirely manual exercise to its current use of software developed for the purpose. Some of the project's achievements, such as antedatings and primary source identification, are highlighted, and challenges, such as unverifiable quotations and sometimes highly convoluted research paths, are described. In addition to this, the article looks at the necessarily systematic nature of quotation handling and the main types of considerations determining methodology (for example, lexicographic, bibliographic and typographic requirements).

Keywords: DICTIONARY, SOUTH AFRICAN ENGLISH, HISTORICAL PRINCIPLES, VERIFICATION, QUOTATION, CITATION, ACCURACY

Opsomming: *Verstewiging van die fondamente: Gedagtes oor die kontro-
lering van aanhalings in 'n historiese woordeboek, met verwysing na A
Dictionary of South African English on Historical Principles.* *A Dictionary of South
African English on Historical Principles* (DSAEHist) het ontstaan uit aanhalingsbewysmateriaal. Dit
bevat net oor 8 000 inskrywings van Suid-Afrikaanse Engels, met ongeveer 45 000 sitate om daardie
woorde te staaf wat as trefwoorde in die woordeboek ingesluit is. Deur die argaiese elektroniese
formaat te gebruik waarin die DSAEHist geset is, het die Woordeboekeenhed van Suid-Afrikaanse

* This article was presented as a paper at the Fourteenth International Conference of the African Association for Lexicography, organised by the Xhosa Department, University of the Western Cape, Bellville, South Africa, 6–8 July 2009.

Engels 'n digitaliseringsproses van die DSAEHist onderneem waartydens die duidelik geword het dat die aanhalings sal baat by 'n volledige hersiening wat die kontrolering van alle aanhalings in hulle oorspronklike bronne behels. Die artikel ondersoek die vordering van die aanhalingskontroleringsprojek vanaf sy begin as 'n algehele handtaak tot sy huidige gebruik van programmatuur wat vir die doel ontwikkel is. 'n Aantal van die projek se suksesse soos vroeëre daterings en primêrebronidentifikasie word uitgelig, en uitdagings soos onkontroleerbare aanhalings en somtyds hoogs verstrengelde navorsingsroetes word beskryf. Hierbenewens kyk die artikel na die noodsaaklik sistematiese aard van aanhalingshantering en die hoofsoorte oorwegings wat metodologie (byvoorbeeld, leksikografiese, bibliografiese en tipografiese vereistes) bepaal.

Sleutelwoorde: WOORDEBOEK, SUID-AFRIKAANSE ENGELS, HISTORIESE BEGINSELS, KONTROLERING, AANHALING, SITAAT, NOUKEURIGHEID

Introduction

A Dictionary of South African English on Historical Principles (1996) was the product of twenty-five years' research, and is rooted in quotation evidence. It contains just over 8 000 South African English entries (including variant forms and unassimilated terms), each of which presupposes sufficient quotation evidence as a prerequisite for inclusion. These quotations are drawn from an extensive card index archive that was built between the late 1960s and early 1990s, providing a pool from which about 45 000 citations were selected to support those words finally included as headwords in the dictionary itself. Because it is a historical dictionary, citations feature prominently in the entry model, providing a set of concise illustrations of their parent headword. These short excerpts are taken from a variety of sources — diaries, letters, newspapers, magazines, novels, reference works, posters, radio, television, speeches, conversations — and not only contain historical and semantic information which it is not possible to represent neatly in a dictionary definition, but also played a key role in the early stages of lexical acquisition. By recording a quotation and indexing it by catchword — that is, the word as it appears in context rather than its canonical form — a potential headword would be noted or existing evidence for it augmented.

Using the legacy electronic format in which *A Dictionary of South African English on Historical Principles* (DSAEHist) was typeset, in 2006, the Dictionary Unit for South African English (DU) embarked on a process of digitisation of the published set of quotations. This involved both conversion into a future-proof, industry-standard format and the development of a searching and updated editing environment. There were two reasons for this digitisation process. The first, not central to this article, was to prepare the entire dictionary dataset for use in modern software. The second, shorter term motivation was to utilise the set of quotations published in DSAEHist as an electronic citations database in its own right. During this conversion process, it became clear that the resulting citations database would benefit from a full review involving the

verification of all quotations against their original sources. This time-consuming process, some stages of which are unavoidably manual, is now over 60% complete, and is on target for final completion by 2011.

This article examines the evolution of the quotation verification project from its beginnings as an entirely manual exercise, to its current use of software developed for the purpose. To provide the context for this process, the past and future purposes of the citations database for the DU are discussed, and some of the project's achievements, such as antedatings and primary source identification, are highlighted. Challenges encountered while trying to verify quotations are also described, such as unverifiable quotations and sometimes highly convoluted research paths. In addition to these, the article will look at the necessarily systematic nature of quotation handling and the main types of considerations determining methodology (for example, lexicographic, bibliographic and typographic requirements). Finally, the impact of the internet on quotation research is also considered, with some discussion of how this may affect the revision process for a new edition of DSAEHist.

Historical dictionaries

Firstly, some comments about diachronic and synchronic dictionaries.

Like the *Oxford English Dictionary* (OED), DSAEHist is a diachronic dictionary, that is, one that traces the origins and development of words through one or more periods in the history of the language. The main purpose of a diachronic dictionary is to trace the development of form and meaning over an extended period of time: the attention of a historical dictionary is directed to the evolution of meaning (Landau 2001: 130). A dictionary such as the OED "not only provides a historical record of the development of meaning of each word, with illustrative quotations and definitions for each sense, it also shows the changes in spelling, the different forms each word assumed during its history" (Landau 2001: 81). As Zgusta (2006: 126) puts it, "Historical dictionaries are ... synchronic descriptions of older historical epochs."

A synchronic dictionary, on the other hand, such as the *South African Concise Oxford Dictionary* (SACOD), focuses on contemporary language, and attempts to represent the lexicon as it exists (or existed) at a particular point in time — generally, but not necessarily, the present — so etymology is less important than in diachronic dictionaries (Landau 2001: 27).

Hartmann and James (1998: 68) describe a historical dictionary as "a type of reference work in which the vocabulary of a language is traced through time". Whereas general dictionaries provide information on the derivation of words — etymology — under the influence of the historical-comparative perspective in linguistics, historical dictionaries trace the formal and semantic changes in the vocabulary of a language throughout its history or in a particular period, documenting the changes in form and meaning of words (the 'curriculum vitae of vocabulary').

The general objectives of a historical dictionary are described by James Murray in the aims of the *Oxford English Dictionary*:

The aim of this Dictionary is (1) to show, with regard to each individual word, when, how, in what shape, and with what signification, it became English, what development of form and meaning it has since received; which of its uses have, in the course of time, become obsolete, and which still survive; what new uses have since arisen and when: (2) to illustrate these facts by a series of quotations ranging from the first known occurrence of the word to the latest, or down to the present day; the word being thus made to exhibit its own history and meaning: and (3) to treat the etymology of each word.

In dictionaries based on historical principles the ordering of homonyms normally displays the historical ordering, and the lemma sign representing the oldest lexical item is presented as the first homonym (Gouws and Prinsloo 2005: 100): the ordering represents the historical development of the various senses (Kastovsky 2000: 119), and there is therefore a challenge for the editors in identifying fine subdivisions of senses (Coleman and McDermott 2004: v). At the Dictionary of Old English project the evidence of citations is used as the primary guide to sense-division (Coleman and McDermott 2004: 138).

In most synchronic dictionaries focusing on the language as it is currently used, the ordering is usually determined by the usage frequency of the words and the word with the highest usage frequency will be presented as the first homonym (Gouws and Prinsloo 2005: 100).

Etymology is an integral part of the work of a historical dictionary, since it illustrates how form and meaning have changed over an extended period of time, often centuries (Landau 2001: 27). The logical requirement of starting with the etymological (or derivational) meaning is much stronger in diachronic dictionaries than in synchronic dictionaries (Zgusta 2006: 29), and DSAEHist uses labels such as obsolete, obsolescent and historical to elucidate this.

Zgusta (2006: 3) points out that "a historical dictionary can offer ... a description of a ... stage in the development of [a] language" or it could be "concerned with the development of a language in the past only, not focusing on or even not considering at all the present state of that language. ... Thus, the expression historical dictionary is used in reference both to period dictionaries, and to diachronic dictionaries, either of them situated on various points or stretches of the flow of time, or development. That many, perhaps even most, dictionaries simultaneously consist of components that belong to different types may go without saying".

Finally, a comment about the OED (<http://www.oed.com/about/>):

As the OED is a historical dictionary, its entry structure is very different from that of a dictionary of current English, in which only present-day senses are covered, and in which the most common meanings or senses are described first. For each word in the OED, the various groupings of senses are dealt with in

chronological order according to the quotation evidence, i.e. the senses with the earliest quotations appear first, and the senses which have developed more recently appear further down the entry. In a complex entry with many strands, the development over time can be seen in a structure with several 'branches'.

Sources of quotations

The choice of quotation sources tells us as much about the values and reading preferences of the editors/readers/lexicographers as they do about the language itself (Brewer 2007: 186) and reinforces, whether consciously or not, a particular set of cultural values (Brewer 2007: 189). Charlotte Brewer describes the OED as a "cultural treasure-house, simultaneously testifying to the longevity of the language, the glories of the English tongue, and the scholarship of the Dictionary's editors" (Brewer 2007: 4). Rowena Fowler points out that it is "possible to read the *Dictionary* [Richardson's *New Dictionary of the English Language* (1836–7)] as a historical document, quarrying it for evidence of early nineteenth-century knowledge, attitudes and mentalities" (Fowler 2004: 54). She adds: "There is a contradiction built into dictionaries claiming to use quotations from the 'best' or 'most distinguished' writers, since these writers may not be typical, characteristic or otherwise exemplary from the lexicographical point of view. The question cuts two ways: can poetic language provide reliable illustration of meaning, and can or should the dictionary provide adequate treatment of literary usage?" (Fowler 2004: 55). Fowler also stresses the lexical dubiousness of drawing on fictional dialogue as evidence of spoken usage (Fowler 2004: 57). However, while fictional dialogue provided printed evidence of use at the time of publication, historical novels should be treated with caution as they might fallaciously bring archaic usages into modern times.

It is clearly important to avoid hapax legomena — words which occur only once in either the written record of a language, the works of an author, or in a single text — and to get a wide spread of sources. In his discussion of Samuel Johnson's *A Dictionary of the English Language*, Robert Burchfield comments that while "the central body of his [Johnson's] illustrative quotations was drawn from the greatest literary, philosophical, and scientific works of the period from Sidney to the Restoration, that is from the 1580s to 1660 ... the quotations form a kind of concordance to the most memorable utterances of his day" (Burchfield and Simpson 2002: 85-86). Richardson had a slightly different approach, "based on the notion that quotations alone, if sufficient in number, could serve to elucidate 'true etymological meaning'", reinforcing his theory that each word had a single immutable meaning (Landau 2001: 77-78).

"In historical dictionaries they (quotations) can show the type of source in which the term is recorded (formal literature, textbooks, magazines, science fiction, etc.). They can show typical (and sometimes eccentric) contexts in which the term has been used, and in a chronological sweep of evidence it is often possible to obtain a telescopic panorama of semantic and structural shift

through which a term has passed over the centuries" (Simpson 2003: 269). While quotations in synchronic and learners' dictionaries can be "invented" examples of usage, invented examples should not be used in historical dictionaries.

With the OED, "Contributors were asked to analyse carefully 'the works of any of the principal writers, extracting all remarkable words, and all passages which contain definitions or explanations' ... The lists also make it clear that writers who use language in markedly idiosyncratic ways were not to be excluded" (Brewer 2007: 124). However, a "reliance on literary quotations is problematic, because it skews the representative character of the sampling" (Brewer 2007: 125).

Mooijaart (2004: 202) lists three important factors that need to be considered when choosing citations:

- (1) the quality of the citations themselves: length, reliability and correctness of the text;
- (2) the accessibility and usability of the electronic citations and the non-linguistic information; and
- (3) the quality of the citations as a corpus: issues of representativeness and diversity.

Mooijaart (2004: 202-203) also points out some of the possible functions of citations:

- As lexicographic evidence
- As "examples of good taste"
- To exemplify meaning
- To exemplify meaning but also to illustrate the variety of grammatical context, connotation and register

A basic principle, however, is that quotations should generally describe and explain linguistic conventions, rather than trying to account for every individual language event (Atkins and Rundell 2008: 48).

In some cases special evidence is taken from reference works such as dictionaries. However, with some exceptions, Mooijaart (2004: 203) feels that, with some exceptions, these are not appropriate sources for illustrative citations as most of them offer invented phrases, or abbreviated phrases with no proper bibliographical documentation of sources.

How DSAE originally collected citations

At the commencement of the DSAEHist project, no electronic corpora were available, and so the process similar to that employed by Murray for the OED was followed: citations were glued on to cards, or handwritten, or typed. Lan-

dau points out that "vast citation files still exist only in paper form because some of the citations cannot be scanned (the type is too faint, or too large, or the citation is handwritten, or there are important handwritten notes attached, or the paper is too fragile, etc.) and because keyboarding would be prohibitively expensive" (Landau 2001: 190). This is certainly the case with DSAEHist: at a rough guesstimate, the DU currently has over 300 000 cards from which quotations for the dictionary were extracted.

Digitisation of DSAEHist

DSAEHist was produced before the advent of dictionary publishing systems, in Wordperfect 5. Before publication these Wordperfect files were sent via FTP to the publisher, the Oxford University Press in Oxford, and then to the US for semi-automated conversion to SGML (a textual format which is the predecessor to XML) to be used for typesetting. This SGML file was also the basis of the new XML version to which DSAEHist was converted in 2005. The quotations were converted, but the rest of the dictionary was only partly converted because of other work pressures, and this therefore became low priority.

Another key benefit of the conversion work, conducted with a good text editor, pattern matching, and custom text-processing utilities developed in-house, is that inconsistencies in the content (including quotations) could be identified and either highlighted for manual attention or fixed globally. The markup used is standards-compliant, transparent markup which is not bound to any particular software: Serna has been most useful so far and so that is what has been used.

The digitisation of the dictionary made it easier to analyse the dictionary and produce statistics such as which authors or newspapers were most frequently quoted and whether there was a good geographic spread across the country (unsurprisingly, there was a heavy bias towards the Eastern Cape, which is where the Dictionary Unit is located), what were the earliest quotations, distribution of oral informants (again, there was a heavy bias towards the Eastern Cape), analysis of chronological range, and so on.

DSAEHist style guide

An extensive style guide was developed by the editors of DSAEHist. This was based largely on the style of OED, but with some modifications based on local requirements. This lengthy document covers matters such as the layout, order and numbering of entries; how titles, authors, page numbers and so on are styled in the dictionary; use and treatment of proprietary names, prefixes and suffixes; parts of speech; labels for subject, register and frequency; etymology; combinations, combinations and compounds; bibliography; abbreviations; spelling (use of *-ize* rather than *-ise* forms); punctuation — and much more. Some of these aspects can be handled by a style sheet within the software.

Database of quotations

It may be useful to distinguish briefly between the digitised version of the dictionary and the database of quotations (the terms 'citation' and 'quotation' are used interchangeably in this article).

As described above, the database existed as SGML quotations, which needed conversion and integration with Serna and the development of a search interface. The conversion was done using in-house text-conversion skills and utilities, as well as commercial text-processing software, in a sort of semi-automated fashion, part manual, part automated. It was a very slow, painstaking process with many safety mechanisms.

The quotations are of course a component of the entire dictionary, and the quotations were simply extracted and treated as a separate database, to be edited on its own. Whether the reviewed quotations are treated separately or as part of DSAEHist is a matter of choice and convenience. The way the markup and styling have been approached means that the typography and styling rules of DSAEHist have for the most part been preserved, unless there has been a conscious decision to change them. Some changes present challenges, such as cases where a quotation date was modified and this was integral to the entry (e.g. the definition referred to it), but these are minor. In a sense the review of quotations was the beginning of the revision of the DSAEHist.

A separate database was created for bibliographical information (see the section **Bibliography** below).

Extracts in the early days of the verification process were less sophisticated, as the database was still in the process of being developed and refined, tags redefined, and so on. So the early extracts did not, for instance, include the headword, or the quotation number, or distinguish between monographs and serials. Now, with the online search system using a web browser, the extracts produced are much clearer and easier to work with: they do not include the tags, they include both old and new quotation IDs, and because of the layout, it is much easier to notice inconsistencies of styling.

Purposes of the citations database

Some of the purposes for which the citations database may be used in future include:

- (1) Integration with a larger database for use in future quotation research
The idea was for the DSAEHist quotations database to be integrated with a new 'intake' (or, as called by the OED, 'incoming') database to provide one searchable, growing set of quotations. New intake will have different emphasis: recent citations for the same words as well as new words, use of internet resources etc. After the verification stage, post-dating quotations for DSAEHist entries could be approached systematically, with use of the internet, etc.

- (2) Integration with DSAEHist editing environment
Quotation handling is an important component of DSAEHist editing, technically as well as lexicographically. The new quotation markup allows software to do automatic sorting and styling of quotations in dictionary entries, generally removing the burden of manual editorial tasks.
- (3) Integration with bibliography
Quotations are OED-style linked with entries in the bibliography entries.
- (4) Semi-automated proofing and statistical checking
The markup makes it possible to identify imbalances in date range, source type coverage and so on with customized queries.
- (5) There are also other possibilities such as linking quotations to library records and internet-published resources, and ranking quotations (e.g. isolating quotations from oral or written informants as opposed to printed quotations, or 'informational' quotations).

Bibliography

The bibliography was converted to a text editor, which is maintained on an ongoing basis as verification of individual titles is completed. The bibliography printed in DSAEHist was of necessity limited — publishers are always concerned about space constraints. But in an electronic file not only is it easier to have more data, it is also possible to record information which may not necessarily be part of the final bibliography available to users of the dictionary, but is extremely useful to the editors as they work. In other words the markup allows custom views (e.g. published and unpublished views) determined by purpose and context.

For example:

- Information regarding the location of titles. It can sometimes take quite a bit of detective work to find out where a title is held, and once it has been tracked down it is useful to record this for future reference. This information includes the Dewey Decimal Classification number as well as the name of the repository.
- Name of publisher. In some cases, there is more than one edition of a title, published in the same year or in different years, with different pagination, and clearly this information is necessary for a reader who wants to look at the original title from which a quotation was taken. This information was not included in the printed bibliography in DSAEHist possibly because of space constraints.
- First names of authors. This is essential for authors such as Margaret Roberts and Michael Roberts, but useful for all authors. If a reader wants

to look up an author called, say, M. Roberts, there may only be one M. Roberts in the DSAEHist bibliography, but many on a library catalogue.

- Additional information about the title, such as the full title as well as the abbreviated 'short' title used in citations. There may be more than one version of the short title: for example, where a diary has been edited and quotations are in some cases taken from the original author, and in other cases from the words of the editor (for example, notes, footnotes or editorial comments):

1835 A. SMITH *Diary* (1940) II. 62 After waiting about ten minutes we were apprised by the exclamations 'Byat! Byat!' that he (i.e. Mazilikatze) was approaching from an inner enclosure.

1940 P.R. KIRBY *Diary of Dr Andrew Smith* II. 72 Bayede, Hail, King.

1940 P.R. KIRBY *Diary of Dr Andrew Smith* II. 72 Bayede, Hail, King. The Matabele royal salute.

The first Kirby quotation is as it appears in DSAEHist, the second version is as it has been finalised after verification, with additional relevant information.

- Notes, such as who has performed the verification, and any other useful information.

Textual accuracy

John Simpson, Chief Editor of the OED, points out in the Preface to the Third Edition (<http://dictionary.oed.com/about/oed3-preface/accuracy.html>): "It is important that quotations represent the texts as originally published, and many thousands of texts are being re-examined and the results fed ... on to the online database."

Simpson also stresses the importance of accurate transcriptions from original texts (Burchfield and Simpson 2002: 180):

It is only by making accurate material available that reliable linguistic analysis can be made. The first edition of the *Oxford English Dictionary*, for example, contains in its illustrative quotations seventeen early instances of the elided form *it's* (= 'it is') from sixteenth-century sources (1575–98). In many cases these quotations were cited from later editions of texts in which the editors had partially modernized the spelling to simplify matters for their readers. But anyone investigating the emergence of forms such as *it's* would be ill-served by relying on this evidence, since most of the examples — especially from the early part of the sixteenth century — in fact do not show any elision.

The Examining the OED project at Hertford College points out that searching the OED is sometimes difficult (<http://oed.hertford.ox.ac.uk/main/content/category/11/43/161/>):

Authors' names, works, and dates of publication are not always consistently rendered in OED. Thus in the first edition of OED, the work by Carlyle on the French Revolution was cited in at least five different forms, while in the second edition, volumes of poems by Auden were usually referred to by their UK titles, but also on occasion by their American titles, with dates different from the UK equivalents. The OED is currently standardising the way in which quotation sources are identified in OED Online, in a huge sub-project of the revision process.

Similar problems have occurred in DSAEHist, and an example of this is given in the section **Examples of improvements** below.

The policy in DSAEHist was to retain the original spelling (but not orthography — the dictionary uses the modern terminal or short 's' rather than what is known as the long, medial or descending 'f'), lineation, and punctuation, except that double quotation marks have been replaced by single quotation marks.

The traditional editorial practice of large historical dictionaries was for selected citations to form complete phrases or full sentences with respect to syntax. Although words or phrases within the citation which were not relevant for the word to be illustrated were often skipped and replaced by ellipses, the basic structure of the sentence was maintained. This implies that all parts of speech are represented, but in the citations they will however have a less complex structure than they have in texts. This was frequently the practice with DSAEHist.

Dating of quotations in DSAEHist

The policy with regard to dating of quotations is based on and similar to that of the OED (see <http://www.oed.com/archive/oed2-preface/gen-main-5.html>). Quotations are normally taken from the earliest available printing of a work; where a later text has been used the publication date is given in parentheses after the title. In DSAEHist, quotations from Pettman (1913), Jeffreys (1964, 1967, 1970) and Swart (1934), three secondary sources who wrote extensively on South African English, were not verified against primary sources but had an attribution in parentheses after the citation. During the present verification process, it has been possible to verify a considerable number of these against primary sources. Where quotations have been taken from more than one edition of a work, the edition used is given in parentheses after the title (this practice is followed where a particular quotation is only present in a later edition).

Verification process

The verification process consists basically of three steps:

- Verification

- Data capture
- Finalisation

(1) Verification

With DSAEHist, as with OED, verification takes place principally by printing out all of the quotations from a particular source — title or author — and comparing these against the appropriate edition of the source text, making amendments where necessary.

(2) Data capture

The next step is for a keyboarder to locate each quotation via an interface programme running under a web browser (Firefox), and to edit it in a specially designed text editing interface (running in Serna). Any changes must be recorded, as well as the initials of the verifier and the data capturer/keyboarder. There is a facility for annotations and queries, which has proved very useful.

(3) Finalisation

The final step is for the original verifier to compare the annotations on the printouts against the updated quotation on the database to ensure that the keyboarder has correctly interpreted handwriting and notes, and correctly captured all changes.

During the verification process, it has been possible to verify a number of quotations against a primary source or an earlier edition — about 2 100 quotations so far.

It has also been possible to antedate a small number of quotations. However, these statistics are not definitive: Because the researchers are looking at quotations from a particular source, rather than quotations for a particular lemma, it would be too time-consuming, and not productive simply for the sake of statistics, to consider each lemma to see whether verifying against an earlier edition would result in an antedating.

Quotation verification system

The quotation verification system (QVS), accessed through a web browser, has options to search in a number of different key fields, such as full text, quotation text, author, title, headword and quotation index. The search interface will compare against any of whole word, whole field, start of word or 'anywhere', and also allows for multiple combinations of search options.

An alternative search option is to filter by verification status. Here one may select from a drop-down box the initials of any of the staff members working on the system, or 'somebody', 'nobody' or 'unfiltered', in each of three fields: Verifier, Capturer and Finaliser. This is useful if, for example, a verifier wants to see which of the verified quotations have been captured and are

waiting to be finalised. It is also useful for extracting statistics, to show how many quotations have been captured and finalised. It does not, obviously, show just how many quotations have been verified, because they are only recorded as being verified at the same time that they are captured, so these two fields will show the same count. The statistics for quotations verified are thus a manual count, maintained on an Excel spreadsheet.

All the verification has been done by one of the editors, Sheila Hicks, with the assistance of a researcher and three student assistants.

Sources for verification

The DSAE has a small library, containing just over 1 600 monographs and about 40 periodical titles. However, we are very fortunate in being able to use the Rhodes University libraries. Some of the quotations were verified against titles in the DSAE library, but a large number came from the Rhodes Library (and the Inter-Library Loans system), the Cory Library for Historical Research, and the National English Literary Museum, as well as the Schönland Herbarium, the Albany Museum, and Rhodes departmental libraries such as Classics, Education, Ichthyology, the International Library of African Music (ILAM), Law, Music and Zoology.

The internet is another useful resource which was not available to the original editors of the dictionary. Sometimes it has been possible to access the 'full view' of a work on Google Books, at other times a 'limited preview' or 'snippet view' has enabled the researcher to request the relevant page numbers through Inter-Library Loans (ILL). In some cases, searching for a particular phrase on Google Books has revealed that the wrong title was cited in DSAE-Hist. Other electronic resources include the Internet Archive, Project Gutenberg, and databases such as JSTOR (available through the Rhodes library system).

When using the ILL system, it has been much easier to locate monographs than periodicals. A monograph title can be requested through ILL, and in cases where the holding library is not prepared to lend the book itself, it is often prepared to make photocopies or scans of the relevant pages, so a list of page numbers is sent (this usually works well, although sometimes additional pages have to be requested because quotations run over from one page to another). However, with periodicals, specific page numbers have to be requested, preferably with the title of the article (which is almost never known) and the author (which is known in only a few instances, because it was often not included in the quotations). For periodicals such as the *Cape Times*, from which the dictionary has 597 quotations, or *S.A. Panorama*, from which the dictionary has 556 quotations, it is clearly not practical to fill out a form for each quotation. For *S.A. Panorama*, the Nelson Mandela Metropolitan University (NMMU) in Port Elizabeth had good holdings of this title, and it was possible to verify the issues which were not held by Cory at NMMU. For other titles, it is necessary to make

several trips to libraries which have extensive holdings such as the National Library of South Africa's 'Places of Legal Deposit' in Cape Town or the Msunduzi Municipal Library in Pietermaritzburg.

What does the QVS system give us?

Because the system was home-grown, it could be tailored to meet our specific requirements, often with modifications being made along the way as the need for a particular facility became apparent. First of all, because the different categories require different information, the quotations were divided into certain basic categories — monographs, serials, ephemera, manuscript, broadcast or informant. For example, a serial could have a volume and part number, or a day and/or month of publication, and the formatting of the quotation in the dictionary is different from that of a monograph, because a monograph will have an author. For example:

1959 E. MPHAHLELE *Down Second Ave* 188 It was later in 1955 that I joined the African National Congress (ANC). I had for some years been torn between it and the All-African Convention (AAC).

whereas a serial may or may not have a 'nested' author, who is the author of the article but not of the whole publication:

1941 C.A. KUSE in *Bantu World* 1 Mar. 4 We will have no constituted cooperation with political organisations like the A.A.C.

1932 *Grocott's Daily Mail* 9 Jan. 3 The flowers which form the subject of these delightful fantasies are Flames, Morning Flowers, Aandbloem.

Facilities available in QVS include:

- Drop-down boxes listing possible values for certain fields, such as 'Content type', where the options are advertisement, caption, cartoon, heading, legend, letter or title. Without such a system, there was plenty of scope for inconsistent values to be used.
- Facility to add a 'minor title', such as the title of a short story. Although this is not printed in DSAEHist, it is particularly useful in, for example, an anthology or collection of short stories, where the minor title can enable the researcher to track down the first date of publication and verify the quotation against the first printing.
- Ability to insert special characters, such as non-breaking spaces (for example, what looks like a space in '65 372', to prevent the numbers being broken over a line-end), or special symbols used in original texts, for example to indicate pronunciation (i.e. not true phonetics) such as

1919 H.H. JOHNSTON *Compar. Study of Bantu & Semi-Bantu Lang.* I. 797 The divergent dialects of ?ōsa, such as Isi-pondw, Isi-baʔa, Feŋgu, &c.

- Facility to add line-break indicators in poetry. Poetry often uses a style of language diverging from that of ordinary speech, and it may be useful to show this. Some readers indicated line-breaks by putting commas at the end of lines and/or beginning new lines with a capital letter, but these may or may not have been there in the original text.
- Facility to add page-break indicators. Where a quotation runs over two pages, this was inconsistently treated in DSAEHist: sometimes the page number on which the quotation begins was given, and sometimes the page number on which the headword occurs. This facility makes it easier to spot such inconsistencies.
- Facility to record column numbers in periodicals. Trying to find a quotation in an old number of a broadsheet newspaper, typically using a very small font, can be quite difficult, but it is much easier if the column number is given.
- Accurate recording of fonts. In DSAEHist, fonts were typically 'normalized', with emphasis such as bold font, small capitals, italics or underlining downgraded or removed. With QVS, it is possible to record fonts accurately, and leave editorial decisions about styling to be made at the time of publication.

Extracting by author makes it easier to spot duplications. For example, here is an excerpt from the entry for **sitkamer**:

1908 F.C. SLATER *Sunburnt South* 12 The room into which he showed me was the *zit-kamer* or sitting-room; it was scrupulously neat and tidy. 1909 F. MASEY in *State* Vol.2 No.7, 67 In the 'zitkamer', .. there stands that commonly met feature the curio cabinet. 1912 F. BANCROFT *Veldt Dwellers* 31 It was smoke-room, bar room, and general zit-kamer combined. 1927 F.C. SLATER in *Outspan* 1 Apr. 5 The room into which he showed me was the zit-kamer or sitting-room; it was scrupulously neat and tidy.

It will be noticed that the quotations dated 1908 and 1927 are the same, but the 1927 one was taken from a secondary source. Here's another example for the same author, from the entry for **place**:

1908 F.C. SLATER *Sunburnt South* 139 He had recently met me at a neighbouring farm and, finding in me an attentive and admiring auditor, had cordially invited me to visit his 'place'. 1929 W.M. MACMILLAN *Bantu, Boer & Briton* 200 (Swart), A great number of the places they deserted are now used as cattle places by proprietors who do not reside upon them but leave them and their cattle in charge of freedmen, Bechuanas and Bushmen. 1931 F.C. SLATER *Secret Veld* 219 He had recently met me at a neighbouring farm, and, finding in me an attentive and admiring auditor, had cordially invited me to visit his 'place'.

Here the incorrect title (and date) had been given for the 1931 quotation.

Examples of the verification of the primary source

At **berg**:

1944 J. MOCKFORD *Here Are S. Africans* 71 The total number of wagons that crossed the Berg during the Great Trek is estimated at more than one thousand.

After verification:

1937 M. NATHAN *Voortrekkers* 177 The total number of wagons that crossed the Berg during the Great Trek is estimated at more than one thousand.

At **sammy**:

1949 J. MOCKFORD *Golden Land* 179 Sammy's baskets balanced on the end of a bamboo pole, like the nests of a weaver bird, go bobbing along the streets and lanes of every township in Natal.

After verification:

1921 C. DAWBARN in *My S. Afr. Yr* 214 Sammy's baskets balanced on the end of a bamboo pole, like the nests of a weaver bird, go bobbing along the streets and lanes of every township in Natal.

In both cases, the primary source was listed in Mockford's books as a footnote, with no page number, so it took some detective work and speed-reading to find the quotation in the primary source.

Here is another example:

blue-coat *n.* *Prison slang.* [tr. Afk. *bloubaadjie*, see BLOUBAADJIE.]

1. A long-term male prisoner serving an indeterminate sentence, usu. after having been declared an habitual criminal; BLOUBAADJIE sense 2 b. See also BLOUROKKIE sense 1.

1949 H.C. BOSMAN *Cold Stone Jug* (1969) 12 In prison the blue-coat occupies a position of some degree of importance. A blue-coat is even higher than a murderer. 1976 V. ROSENBERG *Sunflower* 57 In prison, the murderer, unlike the blue-coat, does not wear a distinctive garb. He is not dressed by the authorities in a way to single him out from the other convicts.

2. An indeterminate prison sentence; BLUE-JACKET; COAT; cf. BAADJIE sense 2.

1949 H.C. BOSMAN *Cold Stone Jug* (1969) 11 He goes back to prison to serve the indeterminate sentence all over again. As they say in prison, he goes back to do his second blue-coat. 1975 *Sunday Times* 15 June 3 If I give myself up, I want to know that I'm not going to get a bluecoat (indeterminate sentence).

Careful examination of Rosenberg's *Sunflower to the Sun*, and some speed-reading of Bosman's *Cold Stone Jug*, reveals that the 1976 quotation should be:

1949 H.C. BOSMAN *Cold Stone Jug* 8 In prison, the murderer, unlike the blue-coat, does not wear a distinctive garb. He is not dressed by the authorities in a way to single him out from the other convicts.

However, this means that both the two quotations given to support the first sense of **blue-coat** are taken from Bosman (as well as one of the two quotations

for the second sense). This is not an adequate spread of sources, and so the editors will have to look carefully at what other quotations are available for this lemma.

A number of quotations in DSAEHist were taken from OED: For example, this one at **sassaby**:

1820 S. DANIELL *Sk. S. Africa* 18 The Sasayby is an Antelope, heretofore not described, found in the Booshwana country.

which in DSAEHist became

1820 S. DANIELL *Sketches S. Afr.* 18 The Sasaybe is an Antelope .. found in the Booshwana country.

It will be noticed that although the title has been changed from the way it was styled in OED ('Sk.' has become 'Sketches'), it still does not conform to DSAEHist styling, which is to include prepositions. Also, as the full title of the work is *Sketches Representing the Native Tribes, Animals, and Scenery of Southern Africa from Drawings Made by the Late Mr. Samuel Daniell, Engraved by William Daniell*, the abbreviation should be 'Sn' not 'S.'. It will also be noticed that the spelling 'Sasayby' in OED has been changed to 'Sasaybe' in DSAEHist. It was possible to obtain the relevant pages of the primary source, and the revised styling of this quotation will now look like this:

1820 S. & W. DANIELL *Sketches .. of Sn Afr.* 18 The Sasayby is an Antelope .. found in the Booshwana country.

Examples of antedating

African Hunting from Natal to the Zambesi, Including Lake Ngami, the Kalahari Desert, etc., from 1852 to 1860 by William Charles Baldwin was first published in 1863. The Cory Library for Historical Research holds the second edition, published in 1894, as well as a facsimile reprint of the first edition. Of the 24 quotations from this work in DSAEHist, nineteen were dated 1863 and five were dated 1894. By checking the first edition for these five, it was possible to antedate the first recording of **inyoka** by three years from 1866 to 1863:

inyoka /in'jɔ(:)ka/ *n.* Also **nhoca**, **nyoka**. [Xhosa and Zulu.] A snake. Also *attrib.*

1866 W.C. HOLDEN *Past & Future* 300 This was the highest kind of sacred inyoka, or 'serpent'. 1891 R. MONTEIRO *Delagoa Bay* 114, I described the noise to Jack in the morning, and he at once said it was made by a 'nhoca' snake as thick as his arm. 1894 W.C. BALDWIN *Afr. Hunting* 112 One of the Amatongas .. gave a most unearthly howl, .. saying that an inyoka snake had bitten him. 1937 [see IZINYANYA]. 1958 R. COLLINS *Impassioned Wind* 60 Inyoka! Snake! Where's the snake?

and of **kaalkop** by 31 years from 1894 to 1863:

2. Special Comb. [...]; **kaalkop** /-kɔp/ [Afk., kop] head, a bald person; also *fig.* and *transf.*; [...]

[...] **1894** W.C. BALDWIN *Afr. Hunting* 301 The bush was very good, a moderate breeze of wind, which I kept always below, but I had great difficulty in getting the bull out from the company of the 'carl kop' (naked head). **1896** R. WALLACE *Farming Indust. of Cape Col.* Kaalkop wheat, a beardless variety well liked by millers. **c1936** S. & E. *Afr. Yr Bk & Guide* 1022 On the silver coins of the Union, the King's head is crowned. The lack of this in British Silver has earned them the name of 'Kaalkop,' bare head. [...]

Here is a quotation for **half-cord** taken from Sir Francis Fleming's *Southern Africa: A Geography and Natural History of the Country, Colonies and Inhabitants from the Cape of Good Hope to Angola*, published in 1856:

1856 F.P. FLEMING *Sn Afr.* 480 *Scomber Capensis, Cuv. and Val. (Halfcord).* .. A large fish measuring from two to three feet.

Fleming's book includes as an appendix Ludwig Pappe's *Synopsis of the Edible Fishes at the Cape of Good Hope*, published in 1853, so it was easy to go to the primary source, which was held at the library of the South African Institute for Aquatic Biodiversity, and verify the quotation there, which resulted in an ante-dating for **half-cord** from 1856 to 1853:

1853 L. PAPPE *Synopsis of Edible Fishes* 23 *Scomber Capensis, Cuv. and Val. (Halfcord).* .. A large fish, measuring from two to three feet.

Examples of mistakes

- OED often gave chapter numbers (in roman numerals) instead of page numbers, and the earlier DSAEHist editors sometimes interpreted these as roman page numbers. For example:

1896 BADEN-POWELL *Matabele Campaign* iv,

in OED was interpreted in DSAEHist as

1896 R.S.S. BADEN-POWELL *Matabele Campaign* p.iv.

This has now been corrected to

1897 R.S.S. BADEN-POWELL *Matabele Campaign* 97.

It will be noticed that the date of the quotation has also been changed: Although the full title of the work is *The Matabele Campaign 1896: Being a Narrative of the Campaign in Suppressing the Native Rising in Matabeleland and Mashonaland*, it was only published in 1897. (DSAEHist has seven other quotations from this title, which are all dated 1897.)

- Page number missing or incorrect in DSAEHist. There could be several reasons for this. Sometimes the keyboarder has transposed the numbers: '298' has been typed as '289'. Sometimes it was because two quotations had been presented on one card, and the keyboarder had captured the same page number for both quotations. Sometimes it has simply been

left out by the keyboarder. Sometimes it has been mistranscribed due to unclear handwriting: '40' has been read as '46'. And sometimes the page number was incorrect due to a page break (as mentioned in the section **What does the QVS system give us?** above).

Sometimes the page number can be found by looking at the cards. However, in some cases there is no page number on the card, or the card is missing. Then some detective work is required. Google Books has proved very useful in tracking down some of these: Even if there is only a snippet view, it may give the page number. In other cases, it is necessary to speed-read or scan the book.

— Volume number missing or wrong.

— Wrong title. This may be a simple typing error, as at **rhenosterbos**:

1941 C.W. DE KIEWIET *Hist. of. S. Afr.* 189 Tumbleweed, prickly pear, rhenosterbos, and jointed cactus invaded the territory of the edible grass and nutritious plants.

where the full stop after 'of' should be deleted.

Or the title may have changed. For example, at **dolf**, the title of the 1st edition was *South West Africa* but the title of the 2nd edition was *Story of Namibia*. Quotations are dated 1976 (1st edition) but the title was taken from the 2nd edition (1978). So

1976 O. LEVINSON *Story of Namibia* 64 Kavango was the most fortunate of the homelands as far as water was concerned ... Timber is a promising industry with the beautiful 'dolf' and 'usivi' (chivi) trees.

should be

1976 O. LEVINSON *S.W. Afr.* 64 Kavango was the most fortunate of the homelands as far as water was concerned ... Timber is a promising industry with the beautiful 'dolf' and 'usivi' (chivi) trees.

Another cause of wrong titles being recorded is that the reader has simply written the author and date on the card, without a title, and the incorrect title has been used. For instance, Keppel-Jones, 1948, p. 60 became:

1948 A. KEPPEL-JONES *When Smuts Goes* 60 The 'Boer with his roer' — his old muzzle-loader — meant security from the old attacks.

This quotation actually comes from Keppel-Jones's *South Africa: A Short History*, published in 1949.

— Wrong font, such as italics instead of roman, or roman instead of italics. One reason for this could be that the original reader underlined the headword on the card in order to highlight it, and of course the conventional proofreading symbol for italics is to underline. Another reason is that it was not possible to capture fonts such as italics, bold, or small

capitals. And a third reason is that bold was used in DSAEHist as a lemma marker, to highlight the first entry in a section of 'special combinations', for example at **protea**:

1962 S. ELIOVSON *Discovering Wild Flowers in Sn Afr.* 50 Often the size of a dinner plate, the flower-head of the **King Protea** is the most spectacular of a large genus.

With the online version, it is possible to indicate this entry differently, so is it now represented thus:

1962 S. ELIOVSON *Discovering Wild Flowers in Sn Afr.* 50 Often the size of a dinner plate, the flower-head of the King Protea is the most spectacular of a large genus.

and the original font preserved. It may be decided in the future to revise the way in which this is indicated: OED online separates quotations for different senses by a horizontal rule, which enables the typestyle of the original source to be preserved.

— Correct punctuation and capitalisation.

— Correct word or spelling, which may or may not change meaning, and may affect the list of variant spellings given for the lemma. Examples include: at **biliary fever** (change 'in' to 'is'), at **looper** (change 'the' to 'they'); at **impala** ('Frangipangi' should be 'Frangipani'); at **Volksraad** ('1938' should be '1838'); at **outspan** ('wanted' should be 'wished'); at **khoisan** ('originated' should be 'diverged'); at **Bantu** ('passionately' should be 'dispassionately', correctly cited at **native**); at **Afrikanerdom** ('unwanted' should be 'unwonted'); at **mealie** ('mealie' should be 'mie-lie'); at **veldskoen** ('feldtschoon' should be 'feldt schoon'); at **liberal** ('negrophile' should be 'negrophilist'); at **saltriem** ('brought' should be 'bought'); at **tessies** ('tessies' should be 'tassies'); at **Uhlanga** ('distinct' should be 'indistinct'); at **werf** ('moved' should be 'removed'); at **pondokkie** ('packing' should be 'sacking'); at **native** ('recommended' should be 'recommenced'); at **stoker** ('decision' should be 'celebration'); and at **rand** ('1918–24 conflict' should be '1914–18 conflict').

— Delete duplicate word:

1898 C. RAE *Malaboch* p.xvii (Jeffreys), The Commissioner sent some of the missionary Kafirs up to to the Hoofstadt with a request for Malaboch to come down as he wanted to take the census of the tribe.

As this was verified in the primary source, it was also possible to remove the details of the secondary source (in this case, Jeffreys), and the spelling 'Hoofstadt' corrected, so that the quotation will now look like this:

1898 C. RAE *Malaboch* p.xvii The Commissioner sent some of the missionary Kafirs up to the Hoofstadt with a request for Malaboch to come down as he wanted to take the census of the tribe.

- Correct author, and add nested author for serials. For example: *A Beginner's Guide to Our Birds* — Jo Oliver was printed as J.O. Oliver.

Examples of improvements

- Nested source information (author and title) added.
- Verify against primary source or earlier edition, especially Pettman, Swart and Jeffreys quotations.
- Take quotations from the same, earlier, edition instead of from several editions:

kaffir: 1908 B. BLACKBURN *Leaven* (1991) 75 The nature of the charge against him was known to every kafir in the jail before he had been there an hour, for news travels fast in kafirdom.

hamba: 1908 D. BLACKBURN *Leaven* 299, I should be only a kitchen boy, as I was in Maritzburg, with the police always waiting to catch me for being out after the 'hamba kyah' bell had rung.

Notice also that the author's initial is incorrect in the first quotation.

- Headword does not appear in quotation:
kill-me-quick: 1948 E. HELLMANN *Rooiyard* 48 At the present time *babaton* comes first in popularity, with *shimeya* or *shimeyani* as close second.

However, the addition of the next sentence resolves this error:

kill-me-quick: 1948 E. HELLMANN *Rooiyard* 48 At the present time *babaton* comes first in popularity, with *shimeya* or *shimeyani* as close second. *Isiqataviki* (kill-me-quick) is made to a much lesser degree.

- Consistency of titles. For example:
1955 T.B. DAVIE *Education & Race Rel.* (Hoernlé Mem. Lecture 1955) 10 Since 1933 .. the Afrikaans-medium universities have disaffiliated themselves [from NUSAS] and have joined a new and separate organization, Die Afrikaanse Studentebond (A.S.B.).

1955 T.B. DAVIE *Educ. & Race Rel.* (Hoernlé Mem. Lecture) 10 As far back as 1924 the students .. created .. a National Union of Students (N.U.S.A.S.) which for many years operated successfully in the interests of students generally.

1955 T.B. DAVIE *Hoernlé Mem. Lecture 1955* 30 The early stages of a restored South Africanism having been prepared in the infant and primary schools and carried forward into the high schools, what can be done to further this aim at the university level?

are now all represented thus:

1955 T.B. DAVIE *Educ. & Race Rel.*

As the full information (that it is The Hoernlé Memorial Lecture 1955) will be given in the enlarged bibliography under the author's name, it is not necessary to include this in the quotation.

- Add additional relevant and useful information. For example, at **Spear of the Nation:**

1964 H.H.W DE VILLIERS *Rivonia* 80 Umkhonto we Sizwe (The Spear of the Nation), was formed under the auspices of the National Liberation Movement.

now reads

1964 H.H.W. DE VILLIERS *Rivonia* 80 Umkhonto we Sizwe (The Spear of the Nation), was formed under the auspices of the National Liberation Movement which included the African National Congress and the South African Communist Party.

(A missing full stop after the author's third initial has also been inserted.)

Unverifiable quotations

It has been impossible to verify all the quotations in DSAEHist. Some of the titles which are not available through any of the lending libraries in South Africa to which the DU has access, or through any of the electronic book repositories, are:

L.P. Greene *The L. Patrick Green Adventure Omnibus* (1925?)

Gwen Westwood *Bright Wilderness* (1970)

Essop et al. *Challenge to 25 Natal Moolvoies* (1969)

The Field (13 Oct 1877 and 9 March 1902)

Suzanna Lynne *Glittering Gold* (1972)

Maximilien Kollisch *The Musselman Population at the Cape of Good Hope* (1867)

E.V.C. *The Promised Land, or, Nine Years (Gold Mining, Hunting, and Volunteering) in the Transvaal* (1884)

Mary Ann Parker *Voyage round the World* (1795)

A number of quotations in DSAEHist were taken from OED, and it has been possible to verify some, but not all, of these in the primary sources. For example, at **sand-mole:**

1850 A. WHITE *Pop. Hist. Mammalia* 232 Another member of this family..is also a native of South Africa: this is the Coast Rat or *Sand-Mole (*Bathyergus maritimus*).

Adam White's *A Popular History of Mammalia: Comprising a Familiar Account of Their Classification and Habits* is not available in South Africa, so OED will have to be relied on as secondary source. The title, however, will be changed to conform to the DSAEHist style, which is to include prepositions, so it will become *Pop. Hist. of Mammalia*.

Another quotation taken from OED for which the primary source is not available in South Africa was (at **imfe**):

1893 WATT *Dict. Econ. Prod. Ind.* VI. iii. 277 This .. is said to be extensively grown in Africa and America, the plant of the former country being the Imphee, and of the latter the Sorgho, which is mainly cultivated on account of sugar.

The abbreviated title does not conform to DSAEHist styling, and the abbreviation 'Prod.' does not appear in the list of abbreviations used in DSAEHist. Although 'Ind.' is given in the list of abbreviations, since this is the only time it is used, the entry will be changed to

1893 G. WATT *Dict. of Econ. Products of India* VI. iii. 277 This .. is said to be extensively grown in Africa and America, the plant of the former country being the Imphee, and of the latter the Sorgho, which is mainly cultivated on account of sugar.

which is rather more reader-friendly. (And the initial of the author has been inserted.)

Conclusion

As time went by, it became more and more evident how necessary it was to do this work, because of the number of quotations that had been, for one reason or another, incorrectly printed in DSAEHist. There will inevitably be some quotations that cannot be tracked down, but the DSAE's quotation verification team feels that it is well on the way to making this a much more accurate representation of the story of South African English.

References

- Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Brewer, Charlotte. 2007. *Treasure-house of the Language: The Living OED*. New Haven Conn./London: Yale University Press.
- Brewer, Charlotte. 2009. *Examining the OED* [online]. <http://oed.hertford.ox.ac.uk/main/> [accessed 27 June 2009].
- Burchfield, Robert, with an Afterword by John Simpson. 2002. *The English Language*. Oxford Language Classics. Oxford/New York: Oxford University Press.
- Coleman, Julie and Anne McDermott (Eds.). 2004. *Historical Dictionaries and Historical Dictionary Research: Papers from the International Conference on Historical Lexicography and Lexicology at the University of Leicester, 2002*. *Lexicographica*. Series Maior 123. Tübingen: Max Niemeyer.
- Fowler, Rowena. 2004. Text and Meaning in Richardson's Dictionary. Coleman, Julie and Anne McDermott (Eds.). 2004: 53-62.
- Gouws, Rufus H. and Danie J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

- Hartmann, R.R.K. and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Jeffreys, M.D.W.** 1964. *Africanderisms*. *Africana Notes and News/Africana Aantekeninge en Nuus* 16(2): 43-95.
- Jeffreys, M.D.W.** 1967. *Africanderisms II*. *Africana Notes and News/Africana Aantekeninge en Nuus* 17(5): 216-220.
- Jeffreys, M.D.W.** 1970. *Africanderisms III*. *Africana Notes and News/Africana Aantekeninge en Nuus* 19(1): 29-41.
- Kastovsky, Dieter.** 2000. Words and Word-Formation: Morphology in *OED*. Mugglestone, Lynda (Ed.). 2000. *Lexicography and the OED. Pioneers in the Untrodden Forest*: 110-125. Oxford/New York: Oxford University Press.
- Landau, Sidney I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.
- Mooijart, Marijke.** 2004. Citations in the *Woordenboek der Nederlandsche Taal*. Coleman, Julie and Anne McDermott (Eds). 2004: 201-211.
- Mugglestone, Lynda.** 2005. *Lost for Words: The Hidden History of the Oxford English Dictionary*. New Haven Conn.: Yale University Press.
- Murray, J.A.H. (Ed.).** 1884. *A New English Dictionary on Historical Principles. Volume 1*. Oxford: Clarendon Press.
- Pettman, Charles.** 1913. *Africanderisms: A Glossary of South African Colloquial Words and Phrases and of Place and Other Names*. London: Longmans, Green.
- Rhodes University Dictionary Unit for South African English (Eds.).** 2002. *South African Concise Oxford Dictionary*. Cape Town: Oxford University Press Southern Africa.
- Silva, Penny and Rhodes University Dictionary Unit for South African English (Eds.).** 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Simpson, John.** 2003. The Production and Use of Occurrence Examples. Van Sterkenburg, P.G.J. (Ed.). 2003. *A Practical Guide to Lexicography*: 260-272. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.
- Simpson, John.** 2009. *Preface to the Third Edition* [online]. <http://dictionary.oed.com/about/oed3-preface/> [accessed 26 May 2009].
- Svensén, Bo.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Translated from the Swedish by John Sykes and Kerstin Schofield. Oxford/New York: Oxford University Press.
- Swart, C.P.** 1934. *Africanderisms: A Supplement to the Rev. Charles Pettman's Glossary of South African Colloquial Words and Phrases and Place and Other Names*. M.A. Thesis. Pretoria: UNISA.
- Zgusta, Ladislav.** 2006. *Lexicography Then and Now: Selected Essays*. Edited by Fredric S.F. Dolezal and Thomas B.I. Creamer. Lexicographica. Series Maior 129. Tübingen: Max Niemeyer.

Access Routes of Internet Finance Dictionaries: Present Solutions and Future Opportunities

Deny Arnos Kwary, *Centre for Lexicography, Aarhus School of Business, Aarhus University, Aarhus, Denmark (deak@asb.dk)*

Abstract: Lexicographers working on finance dictionaries have not properly exploited the fast development of the Internet. This is revealed by the poor utilization of the access routes found in current Internet finance dictionaries. Consequently, users cannot effectively and efficiently find answers to their lexicographical problems. Firstly, there are Internet finance dictionaries which lack technological features and can simply be called meagre Internet dictionaries. Secondly, there are Internet finance dictionaries which utilize so many technological features that users incur high lexicographical information costs. In order to create better access routes, a sound theoretical foundation has to be established. This article shows that the implementation of the modern theory of lexicographical functions, which focuses on the users, results in a better design for future Internet finance dictionaries. With the proper theoretical basis, lexicographers will be able to create state-of-the-art dictionaries that can provide effective and efficient solutions to lexicographical problems.

Keywords: DICTIONARY, PAPER DICTIONARY, INTERNET DICTIONARY, LSP DICTIONARY, FINANCE DICTIONARY, FINANCIAL TERMS, LEXICOGRAPHY, LEXICOGRAPHICAL FUNCTIONS, TEXT RECEPTION, USER NEEDS, USER SITUATION, ACCESS ROUTES, SEARCH OPTIONS, LEXICOGRAPHICAL INFORMATION COSTS

Opsomming: Toegangsroetes van Internet- finansiële woordeboeke: Huidige oplossings en toekomstige geleentheid. Leksikograwe wat aan finansiële woordeboeke werk, het nie die vinnige ontwikkeling van die Internet behoorlik benut nie. Dit blyk uit die gebrekkige aanwending van die toegangsroetes wat in huidige Internet- finansiële woordeboeke aangetref word. Gevolglik kan gebruikers nie doeltreffend en doelmatig antwoorde kry op hul leksikografiese probleme nie. Eerstens is daar Internet- finansiële woordeboeke waarin tegnologiese kenmerke ontbreek en wat gewoon karige Internetwoordeboeke genoem kan word. Tweedens is daar Internet- finansiële woordeboeke wat so baie tegnologiese kenmerke aanwend dat gebruikers hoë leksikografiese inligtingskoste oploop. Om beter toegangsroetes te skep, moet 'n deeglike teoretiese grondslag gevestig word. Hierdie artikel toon dat die toepassing van die moderne teorie van leksikografiese funksies wat op die gebruikers fokus, lei tot 'n beter ontwerp vir toekomstige Internet- finansiële woordeboeke. Met die gepaste teoretiese basis, sal leksikograwe in staat wees om woordeboeke te skep wat die jongste tegnologie gebruik om doeltreffende en doelmatige oplossings vir leksikografiese probleme te bied.

Sleutelwoorde: WOORDEBOEK, PAPIERWOORDEBOEK, INTERNETWOORDEBOEK, TSD-WOORDEBOEK, FINANSIËLE WOORDEBOEK, FINANSIËLE TERME, LEKSIKOGRAFIE, LEKSIKOGRAFIESE FUNKSIES, TEKSONTVANGS, GEBRUIKERSBEHOEFTE, GEBRUIKERSITUASIE, TOEGANGSROETES, SOEKKEUSES, LEKSIKOGRAFIESE INLIGTINGSKOSTE

1. Introduction

Recent advances in Internet technology offer numerous features which can be incorporated into Internet dictionaries. However, these advances have not been properly embraced by lexicographers. Nielsen and Mourier (2005: 110) noted that some Internet dictionaries are based on printed dictionaries and not developed directly for the electronic medium. They therefore do not take advantage of the features offered by Internet technology. Some other Internet dictionaries may have been developed directly for the electronic medium, but the lack of a sound theory which focuses on user needs has made these dictionaries incapable of thoroughly and swiftly satisfying the needs of the users.

A clear example of the poor response to the technology and user needs is the utilization of access routes. The access routes offered by the current Internet Finance Dictionaries (IFDs) are either ineffective or inefficient in helping users find what they are looking for. The search options, which should provide the quickest access route to reach a result, are not efficiently designed. In addition, the results given are not effective in helping the users to understand the terms they are looking for. Consequently, a better concept for the access routes of Internet finance dictionaries is needed, and, as De Schryver (2003: 188) states, a sound underlying theory is obviously needed in creating future electronic or Internet dictionaries.

2. Theoretical Foundation

In order to improve the quality of IFDs, users should become the point of departure in creating the dictionary. This is in line with Bogaards (1999: 118) who urges lexicographers 'to adopt more often and more systematically the viewpoint of the learner' or the dictionary users. In addition, sound theoretical foundations should also be taken into account. In this case, the modern theory of lexicographical functions proposed by Tarp (2008) is applied as it focuses on the users. The concept of lexicographical information costs proposed by Nielsen (2008) is also employed so that the solutions are based not only on the availability of the technological features but also on the calculation of the lexicographical information costs.

The concept of lexicographical information costs was initially proposed by Nielsen (1999). In a later article, Nielsen (2008) divides lexicographical information costs into two distinct types: search-related information costs (SRIC) and comprehension-related information costs (CRIC). SRIC are 'the costs (i.e. efforts) related to the look-up acts users have to perform when consulting a dictionary in order to gain the data for which they are searching', whereas CRIC are 'the costs (i.e. efforts) related to the user's ability to understand and interpret the data presented in the dictionary' (Nielsen 2008: 173-174). Both the SRIC and CRIC are indispensable, as shown in the following sections, for examining the current Internet finance dictionaries and for proposing better search options.

3. Examining Present Solutions of Current Internet Finance Dictionaries

The current Internet finance dictionaries contain disproportionate numbers of technological features; some of them use too few features while others provide too many features. Therefore, they can generally be categorized into two groups: meagre Internet dictionaries and extravagant Internet dictionaries. Meagre Internet dictionaries are Internet dictionaries that utilize so few technological features that they result in poor quality technological capabilities, whereas extravagant Internet dictionaries are those that use technological features in such a careless way that these become wasteful.

3.1 Meagre Internet Dictionaries

There are two typical examples of Internet finance dictionaries which can be categorized as meagre Internet dictionaries. The first one is the Internet version of the paper dictionary entitled *The Language of Money* by Carew (1996). The other is the Internet version of the paper dictionary entitled *The New York Times Dictionary of Money and Investing* by Morgenson and Harvey (2002). The Internet version of *The Language of Money* can be found in the website of ANZ Bank. It appears to be part of the ANZ financial literacy program, called MoneyMinded, which comprises educational resources for people in need of financial education. The Internet version of *The New York Times Dictionary of Money and Investing* is available at Duke University, where one of the authors, Harvey, is Professor of International Business.

Figure 1: Web Page of the Internet Version of *The Language of Money*



When opening the website of *The Language of Money*, users are presented with a picture of the cover of the printed dictionary and an alphabetical list of the dictionary entries, as shown in Figure 1. The search text-box (marked 'Search for')

and the search button (marked 'Go') on the right-hand top are not for searching the terms available in the dictionary, but for searching through the website of ANZ bank. Therefore, when using this dictionary, users cannot type in the term they are looking for.

When users are presented with a web page like this, their look-up acts have to be similar to those for paper dictionaries. For example, if users want to find the definition of the term *debenture*, they need to click the letter D, and then a new web page will pop up, listing all the entries under the letter D. This is shown in Figure 2. Owing to the standard size of the computer screen, only the first few entries under the letter D can be shown on the screen. Therefore, users have to scroll down the screen to find the term *debenture* they are looking for, because it is not on top of the list.

Figure 2: Web Page of Entries under the Letter D of *The Language of Money*



After having scrolled down the screen, users finally find the term they are looking for. However, the look-up acts have not finished, because the definition presented for the term is not complete. It seems that the designer of the Internet version of this dictionary wanted to shorten the length of the page, so only the first two lines of the definition are presented for each entry word. Dictionary users still have to click the entry word to see the complete definition. Consequently, users need to perform at least four look-up acts to reach the dictionary article they are looking for. This certainly takes too much time. In addition, there is no quick link from a term found in the definition to its dictionary entry. Users have to start all over again to find the definition of a term used within the definition and which they do not understand. This dictionary barely utilizes available technological features, so it can simply be called a meagre Internet dictionary.

The second dictionary, *The New York Times Dictionary of Money and Investing*, provides quick links from the terms found in the definitions to their dictionary articles. However, the access routes to the dictionary articles are still similar to the printed version of the dictionary. Therefore, it can be categorized as a meagre Internet dictionary. As Figure 3 shows, users cannot type in the term they are looking for, because the web page has no text-box and search button. The web page only shows the cover of the printed dictionary and an alphabetical list of the dictionary entries.

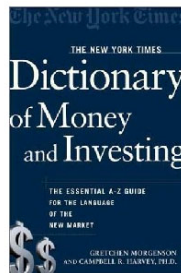
Figure 3: Web Page of the Internet Version of *The New York Times Dictionary of Money and Investing*

[Campbell R. Harvey's Hypertextual Finance Glossary](#)

Over 8,000 Entries and 18,000 Hyperlinks

The largest financial glossary on the Internet

Order the popular hard copy version coauthored with the 2002 Pulitzer Prize winner for financial writing, Gretchen Morgenson of the New York Times.



[Order via Amazon](#)
[Order via Barnes and Noble](#)

Go to definitions organized by letter

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

To reach a dictionary article, users will have to follow access routes that are similar to those for paper dictionaries. Using the previous example to find the definition of the term *debenture*, users must click on the letter D and another web page will pop up, as shown in Figure 4. The web page presents all the entries for the letter D, but the size of the computer screen prevents users from directly seeing the entry for *debenture*. Users have to scroll down the screen to find the term they are searching for.

The only difference between this dictionary and the first is the presentation of the dictionary articles. In the first dictionary, only the first two lines of the definition are presented for each dictionary entry, so users still have to click the entry to see the complete definition. In the second dictionary, users are directly presented with the complete definition. Therefore, instead of perform-

ing four look-up acts, it will only take three look-up acts to reach the dictionary article users are looking for.

Figure 4: Web Page of Entries under the Letter D of *The New York Times Dictionary of Money and Investing*

Campbell R. Harvey's Hypertextual Finance Glossary

Copyright © 2009. All Worldwide Rights Reserved. Do not reproduce without explicit permission.

*Order the popular hard copy version coauthored with the
2002 Pulitzer Prize winner for financial writing,
Gretchen Morgenson of the New York Times.*

[Order via Amazon](#)

[Order via Barnes and Noble](#)

Featured Links

ABCDEFGHIJKLMN OPQRSTUVWXYZ

D

Fifth letter of a [NASDAQ](#) stock symbol specifying that it is a new issue, such as the result of a reverse split.

D/A

See: [Documents Against Acceptance](#)

DCF

See: [Discounted Cash Flows](#)

DSCR

See: [Debt-service coverage ratio](#)

DDM

The [ISO 4217](#) currency code for former East Germany Ostmark.

As becomes clear from the above description, these two Internet dictionaries can undoubtedly be called meagre Internet dictionaries. As shown in the access process to the dictionary entries, both of them use only very few technological features for the Internet versions of the dictionaries. This results in high search-related information costs, as users have to perform at least three look-up acts when consulting these dictionaries in order to arrive at the data they are searching for.


3.2 Extravagant Internet Dictionaries

As explained above, there are Internet finance dictionaries which utilize too few technological features. In contrast, other Internet finance dictionaries use too many technological features, which are carelessly presented. These finance dictionaries can be called extravagant Internet dictionaries. Two examples of this type of Internet finance dictionaries are www.investopedia.com/dictionary/default.asp, which is provided by Forbes Digital Company, and www.financial-dictionary.thefreedictionary.com, which is managed by Farlex.

When opening the website, www.investopedia.com/dictionary/default.asp, users are presented with three options to search for a term, i.e. 'By Keyword' (a text-box with a search button), 'By Alphabet', and 'By Category'. There is also information about 'Top 10 Searched Terms' and 'Recently Added Terms' (see Figure 5).

Figure 5: Web Page of www.investopedia.com/dictionary/default.asp

Financial Dictionary

 **Term of the Day**
Forex Trading Strategy - Forex - Investopedia

*The EURO is **EASY** to trade.*
Forex trading involves significant risk of loss. See How **CLICK HERE**

Dictionary Search Tools

By Keyword (Search our database for over 7000 terms)

Enter a Term Here

By category

- [Active Trading](#)
- [Banking](#)
- [Bonds](#)
- [Brokers](#)
- [Buzz Words](#)
- [Currencies](#)
- [Economics](#)
- [Financial Theory](#)
- [Foreign Exchange](#)
- [Formulas](#)
- [Fundamental Analysis \(Accounting\)](#)
- [Futures](#)
- [General Finance & Investing](#)
- [Hedge Funds](#)
- [Insurance](#)
- [Laws & Regulations](#)
- [Mutual Funds](#)
- [Options](#)
- [Personal Finance](#)
- [Real Estate & Property](#)
- [Retirement Planning](#)
- [Statistics](#)
- [Stocks](#)
- [Taxes](#)
- [Technical Analysis](#)
- [Venture Capital and IPOs](#)

By Alphabet

| [A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

Top 10 Searched Terms

- 1) [EBITDA](#)
- 2) [CAGR](#)
- 3) [P/E Ratio](#)
- 4) [Bo Derek](#)
- 5) [GAAP](#)
- 6) [EPS](#)
- 7) [CFA](#)
- 8) [Warren Buffett](#)
- 9) [Short Selling](#)
- 10) [Stop Loss Order](#)

Recently Added Terms

- [Islamic Financial Services Board - IFSB](#)
- [Sharia](#)
- [Acquiree](#)
- [Acquirer](#)
- [At Sight](#)
- [Lock In Profits](#)
- [Locked-In Interest Rate](#)

From these menus, it can reasonably be assumed that the 'By Keyword' menu is the one most commonly used, because Internet dictionary users, who are also

Internet users, are used to typing a word into a text-box when searching the Internet for information. However, when this 'By Keyword' menu is used, it will become apparent that the lexicographical information costs incurred are high. For instance, after typing the term *debenture* and clicking the search button, users are directed to another web page (see Figure 6) which presents a list of dictionary terms (links) related to the term *debenture*. Users need to click one of the links to see the relevant definition. Therefore, the search-related information costs (SRIC) are high because users have to perform two look-up acts instead of just one to find a definition. The technology enables the website to present results of related terms, but it is not the first information users need when consulting a dictionary to search for definitions of the terms.

Figure 6: Search Results from www.investopedia.com/dictionary/default.asp



In the definition itself, users are still faced with high comprehension-related information costs (CRIC), because of the lengthy article and the difficult words without any cross-references to other dictionary articles. The definition of the term *debenture* given in this dictionary reads as follows:

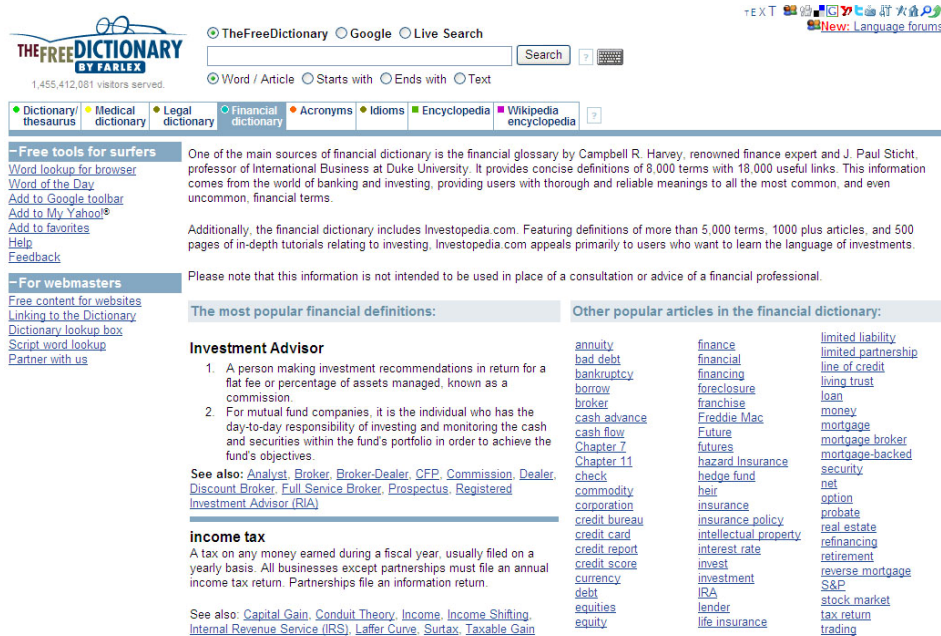
Debenture

A type of debt instrument that is not secured by physical asset or collateral. Debentures are backed only by the general creditworthiness and reputation of the issuer. Both corporations and governments frequently issue this type of bond in order to secure capital. Like other types of bonds, debentures are documented in an indenture.

Terms in the definition, like *collateral* and *indenture*, may cause a comprehension problem to users. In addition, since there are no cross-reference links to

these terms, users have to type these terms again in the text-box to find the definitions instead of simply clicking the terms. Consequently, both the CRIC and SRIC incurred are high. On the one hand, this dictionary has used the Internet technological features in the design by providing three options to search for a term and also information about 'Top 10 Searched Terms' and 'Recently Added Terms'. On the other hand, the technological features used are not really helpful for users, only leading to high lexicographical information costs.

Figure 7: Web Page of www.financial-dictionary.thefreedictionary.com



The second dictionary, www.financial-dictionary.thefreedictionary.com, also contains some technological features. However, the lexicographical information costs incurred are still high. The first web page of this dictionary, as shown in Figure 7, presents a text-box with four radio-buttons which enable users to search for 'Word/ Article', 'Starts with', 'Ends with', and 'Text'. There are also two menus: 'The most popular financial definitions' and 'Other popular articles in the financial dictionary'. These two menus, however, are not very useful for dictionary users. It is unlikely that anyone will open a dictionary in order to find the most popular definitions or popular articles. Moreover, these additional menus may distract dictionary users from their main purpose of consulting the dictionary, i.e. finding the definition of a term.

On the right-hand side of the text-box, there is an icon which looks like a small keyboard. When users click this icon, a virtual keyboard will pop up and users can click on the buttons on the virtual keyboard to type a term into the

text-box. For instance, when users want to search for the term *debenture*, they can either click on the text-box and type the word or click the keyboard icon and click each letter. A test reveals that it takes more time to click each letter from the virtual keyboard than to type the word directly. Therefore, this technological feature is not really useful.

After a term, e.g. *debenture*, has been searched for, it becomes apparent that this dictionary utilizes the technological features excessively, which results in high lexicographical information costs. The search result for *debenture* (Figure 8) shows four definitions, taken from four different dictionaries. This results in high CRIC because users will firstly have to read all four definitions and secondly will not know how to choose the best one. Therefore, the technology, which enables the website to present results from several dictionaries at once only confuses users.

Figure 8: Search Results from www.financial-dictionary.thefreedictionary.com

THE FREE DICTIONARY BY FARLEX
1,455,416,082 visitors served.

TheFreeDictionary Google Live Search

Debenture

Word / Article Starts with Ends with Text

Dictionary/thesaurus Medical dictionary Legal dictionary Financial dictionary Acronyms Idioms Encyclopedia Wikipedia encyclopedia

Debenture Also found in: [Dictionary/thesaurus](#), [Legal](#), [Acronyms](#), [Encyclopedia](#), [Wikipedia](#), [Hutchinson](#)

[Treasury Bonds](#) Find Providers of Treasury Bonds. Your Business Solution Business.com
www.business.com

[450%+ Investment Returns](#) High Yield Alternative Investments. \$10,000 Minimum Investment
www.seismaresearch.com

Debenture
Any [debt obligation](#) backed strictly by the [borrower's](#) integrity, e.g. an [unsecured bond](#). A debenture is documented in an [indenture](#).

Copyright © 2004, [Campbell R. Harvey](#). All Rights Reserved.

debenture
A corporate bond that is not secured by specific property. In the event that the issuer is liquidated, the holder of a debenture becomes a general creditor and therefore is less likely than the secured creditors to recover in full. Because of their high risk factor, debentures pay higher rates of interest than secured debt of the same issuer. See also [subordinated debenture](#).

Wall Street Words: An A to Z Guide to Investment Terms for Today's Investor by David L. Scott. Copyright © 2003 by Houghton Mifflin Company. Published by Houghton Mifflin Company. All rights reserved.

Debenture
A [debt security](#), issued by a government or large company, that is not secured by an [asset](#) or [lien](#), but rather by the all [issuer's](#) assets not otherwise secured. That is, a debenture carries no collateral and is considered unsecured; in case of [bankruptcy](#), the debenture holder is considered a [general creditor](#). A debenture can be traded, and the term is often interchangeable with a [bond](#). Debentures issued by governments are considered risk-free. See also: [Treasury security](#).

Farlex Financial Dictionary. © 2009 Farlex, Inc. All Rights Reserved.

Debenture. A debenture is an unsecured bond. Most bonds issued by corporations are debentures, which are backed by their reputation rather than by any collateral, such as the company's buildings or its inventory.

Although debentures sound riskier than secured bonds, they aren't when they're issued by well-established companies with good credit ratings.

Dictionary of Financial Terms. Copyright © 2008 [Lightbulb Press, Inc.](#) All Rights Reserved.

From the discussion above, it is clear that the search options provided in these dictionaries do not really help users find what they are looking for in the shortest time and in the most comprehensible way. Therefore, it is necessary in the future to create Internet finance dictionaries with better lexicographical solutions so that the needs of dictionary users can be properly satisfied.

4. Future Opportunities for Internet Finance Dictionaries

In order to create a dictionary which can satisfy the needs of users properly, it is necessary to focus on its function and the intended users. By focusing on a particular function, lexicographers will be able to provide better solutions to satisfy the predetermined function. This is not to say that IFDs must only have one function, but it is better to work on one function at a time during the design and development phases. In addition, focusing on a particular user group may also result in better IFDs, because the lexicographical solutions provided can be tailored to the competences of the particular user group.

4.1 Focusing on a Particular Function

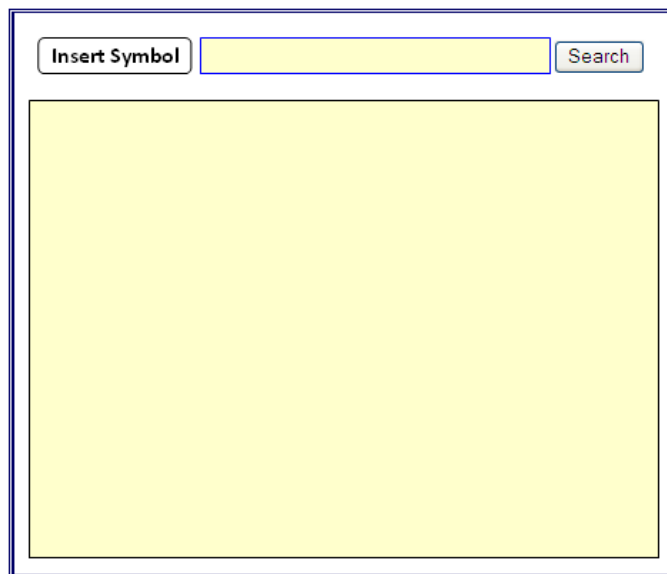
In focusing on a particular function, the modern theory of lexicographical functions initially proposed by Tarp (1994) can be used as it emphasizes the functions of a dictionary as the point of departure in creating the dictionary. A lexicographical function is 'the efforts and ability of a dictionary to provide answers to the complex needs arising in a user in a specific usage situation' (Tarp 1998: 123). As the name of the theory suggests, the first step in creating a dictionary concept is to determine the function of the dictionary for the user group. For communicative oriented functions, according to Bergholtz and Tarp (2003: 176), a dictionary can be used to assist users in solving problems related to:

- (a) text reception in the native language,
- (b) text production in the native language,
- (c) text reception in a foreign language,
- (d) text production in a foreign language,
- (e) translation of texts from the native language into a foreign language, and
- (f) translation of texts from a foreign language into the native language.

In the case of dictionaries of English financial terms for non-native speakers of English, it is the third out of these six functions which is mostly sought. It is more common for learners, who are LSP dictionary users, to read than to write or translate L2 LSP texts. Because their learning materials usually comprise L2

texts, they will need to read and understand the L2 text. However, they will not be asked to translate the texts, neither to write the assignments in L2. Therefore, the most important function in this case is text reception in a foreign language. With this function concept in mind, a possible design for an Internet finance dictionary can be created, as shown in Figure 9.

Figure 9: Design of an Internet Finance Dictionary Focusing on a Particular Function



The Internet finance dictionary has four items, i.e. a text-box, a search button, a search result box, and an insert symbol icon. The text-box is used to type the term, and the search button is used to execute the search. Then, the result is shown in the search result box on the web page. Users should not be directed to another web page or to a list of related terms as in www.investopedia.com/dictionary/default.asp, because users do not need such information. What they need is the dictionary article or the definition of the term. In addition, the dictionary should only present a single definition rather than four definitions from four different dictionaries as in www.financial-dictionary.thefreedictionary.com, so that users will not spend too much time reading the different definitions and be confused about choosing the right one.

In a text reception situation, users encounter a term which they do not understand and they consult the dictionary to find the meaning of the term. They exactly know the spelling of the term as they are reading it, so they can just type the term into the text-box. A problem in typing the term occurs when the term is not in Roman spelling, for instance, terms like β ('market risk') and μ ('population mean') which are quite common in financial texts. It is impossible

for users to type them as these characters are not on the computer keyboard. Therefore, an insert symbol icon should be provided. When users click this icon, a virtual keyboard will pop up, listing the common symbols found in financial texts. Users can click the relevant symbol and it will be directly inserted into the text-box, so that users can search for its meaning.

Other options, such as 'Top 10 Searched Terms', 'Recently Added Terms', 'The most popular financial definitions' and 'Other popular articles in the financial dictionary', offered in the dictionaries examined in the previous section, are not only unnecessary but also clutter the computer screen and distract users from finding the solutions to their problems in the fastest way. The simple Internet dictionary design shown in Figure 9 better provides the solutions to text reception problems than the complicated designs shown in Figures 5 and 7, since in this simple design, the lexicographical information costs incurred are lower. The SRIC is lower because users are shown the result directly after they have clicked the Search Button, and the CRIC is also lower because the results are not based on a compilation of definitions from several dictionaries.

4.2 Focusing on a Particular User Group

A further step which can be taken after determining the function of the dictionary considers the main elements included in the function concept. There are four main elements of lexicographical functions: types of potential user, user situation, user need, and assistance to meet the needs (Tarp 2008: 43). Determining those elements will result in a more focused dictionary with better solutions to lexicographical problems of a particular user group. To see how these elements are implemented in a dictionary project, this article takes the example of the on-going project on the *English Dictionary of Finance for Indonesian CFA (Chartered Financial Analysts) Candidates*. However, the solutions proposed on the access routes may also be applicable to other LSP dictionaries with similar types of users.

For this dictionary, the potential users are Indonesians who have completed undergraduate programs from a faculty of economics or a school of business and are preparing to take the CFA examinations. To draw up a more specific profile of this intended user group, the following eight characteristics proposed by Bergenholtz and Nielsen (2006: 285-286) can be considered:

- (a) Which language is their native language?
- (b) At what level do they master their native language?
- (c) At what level do they master a foreign language?
- (d) How extensive is their experience in translating between the languages in question?
- (e) What is the level of their general cultural and encyclopaedic knowledge?

- (f) At what level do they master the special subject field in question?
- (g) At what level do they master the corresponding LSP in their native language?
- (h) At what level do they master the corresponding LSP in the foreign language?

Since the function of the dictionary is to assist in text reception, characteristics (d) and (e) can be disregarded, because they refer to translation experience and a cognitive function respectively. Consequently, for this dictionary, the user profile is as follows:

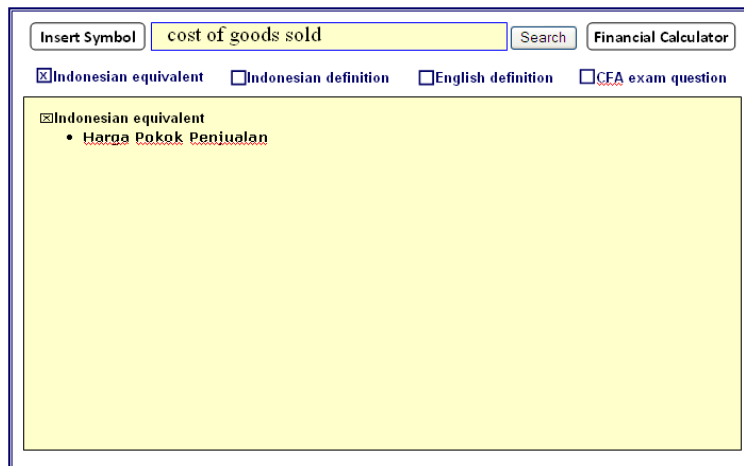
- (a) Their native language is Indonesian, because they are Indonesians.
- (b) They master Indonesian at an advanced level, because they are students who have studied up to undergraduate level in Indonesia.
- (c) They master English at an intermediate level, because they have studied English for at least seven years.
- (d) They are semi-experts in finance, because they have completed an undergraduate program where they have taken several courses related to finance.
- (e) They master Indonesian financial terminology at an intermediate level, because they have learned some financial terms in Indonesian during their undergraduate program.
- (f) They master English financial terminology at a basic level, because they rarely encountered English financial terms during their undergraduate program.

The user situation relevant for this dictionary is reading American financial texts, mainly from the CFA Institute. Therefore, the primary need of the intended user group is help to understand the texts they read, especially when they find terms of which they do not know the meaning. Since they are reading the texts in the framework of preparing for the CFA examinations, the secondary need of these users is help to answer the CFA examination questions. After having determined the potential users, user situation, and user needs, assistance to satisfy the needs can be provided, as shown in Figure 10.

The design features nine items: a text-box, a search button, a search result box, an insert symbol icon, a financial calculator icon, and four checkboxes. The functions of the text-box, the search button, the search result box and the insert symbol icon are the same as those explained for Figure 9. In addition, four further search options are provided. These search options are designed with checkboxes so that users can, if they like, choose more than one option. The first checkbox is for 'Indonesian equivalent' which is also the default option.

When users click the search button directly after having typed in the English term, only the Indonesian equivalent of the English term is presented in the search result box. According to Bergenholtz and Johnsen (2007: 10), 'if a user wants to understand a word, showing only the equivalent is adequate'. For this reason the Indonesian equivalent has been set as the default option. Moreover, since the users are semi-experts in finance, master Indonesian financial terminology at an intermediate level, and master English financial terminology at a basic level, providing the Indonesian equivalent is the best option as it will lead to the lowest CRIC in most cases.

Figure 10: Design of an Internet Finance Dictionary Focusing on a Particular User Group



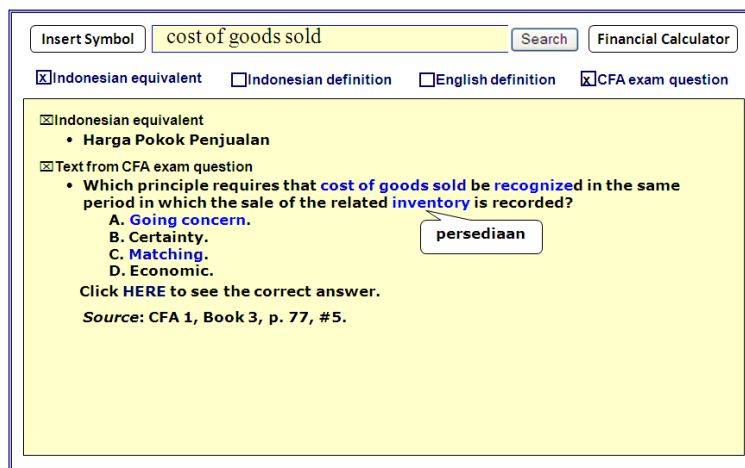
The second checkbox is for 'Indonesian definition', so when users tick this box, the Indonesian definition of the English term will be presented in the search result box. This option is useful when the Indonesian equivalents are only a transfer of the English spelling to the Indonesian spelling patterns. For example, the Indonesian equivalent of the English term *debenture* is *debentur*. Both of these are incomprehensible to the users, so they need a further explanation which can be provided from the checkbox for 'Indonesian definition'. It is also sensible to put this checkbox as the second option before the checkbox for 'English definition', because the users have mastered Indonesian at an advanced level, and English only at an intermediate level. The CRIC incurred will be lower for 'Indonesian definition' than for 'English definition'.

In some situations, there are users in this group who may choose the 'English definition' because they would like to see how the term is defined in English or because they want to practise their English. This is sensible because they are also in the situation of preparing to take the CFA examination which is in English. Hence, a checkbox for 'English definition' is also given in this dic-

tionary. Furthermore, considering that "financial jargon uses a great many acronyms and abbreviations" (Carew 1996: vii), the option 'English definition' can also be the place to put the extension of acronyms and abbreviations, together with their definitions as necessary.

The fourth option, a checkbox for 'CFA exam question' is provided together with a financial calculator icon, which will be useful when users want to see how the term is used in the examination and want to answer a sample of examination questions. However, since the users have mastered English at an intermediate level and have mastered English financial terminology at a basic level, it is necessary to provide a quick reference to the difficult words in the 'English definition' and 'CFA exam question'. For instance, as can be seen in Figure 11, when the word or the term in the search result box is clicked, a 'call out' will pop up and show the 'Indonesian equivalent' of the term. Therefore, users do not need to type the term again in the text-box to find the definition. The 'call out' option provides access with low SRIC.

Figure 11: Internet Finance Dictionary Showing Two Search Results



There are two checkboxes chosen in Figure 11, i.e. 'Indonesian equivalent' and 'CFA exam question'. This is an example of a situation where a user has a problem in comprehending a term like *cost of goods sold*. The user starts by typing the term into the text box and clicking the Search button or pressing the Enter button in the keyboard, and he is directly presented with the Indonesian equivalent of the term. Then, if he wants to know how this term is used in an examination, he clicks the checkbox 'CFA exam question' and an example of an examination question is immediately shown below the Indonesian equivalent.

Although there are two results shown in Figure 11 for the term *cost of goods sold*, it does not mean that the CRIC will be high. In this case, the capability of the technology to present more than one result is utilized well. It is different from www.financial-dictionary.thefreedictionary.com (Figure 7), which

presents definitions from several dictionaries at once. The first difference is that in the design in Figure 7, users cannot choose whether they want to see just one definition or more than one, whereas the design in Figure 11 allows users to choose the data to be presented. The second difference is that the design in Figure 7 presents several definitions and users may not know which one they should follow and which ones are redundant, whereas the design in Figure 11 presents two sets of data which are mutually beneficial, since the first one is the equivalent, which enables users to understand the term, and the second one is an example from the examination which shows the term in actual use. Consequently, this dictionary design meets both the primary and the secondary needs of the users.

The above discussion clearly illustrates that, by knowing the users well, lexicographers are able to create dictionaries that give users what they need (Atkins and Rundell 2008: 28). More importantly, the data presented at any given time or look up fits in with the needs of a particular user in a given user situation (Andersen and Nielsen 2009: 360). All options presented in this proposed dictionary design are tailored to the needs of the users so that they can find the best solutions to their lexicographical problems. Such a dictionary can be called a proper Internet dictionary, because it takes advantage of technological features to provide a lexicographical solution in such a well-planned way that users can solve their lexicographical problems efficiently and effectively. If this proper Internet dictionary is to be called a real Internet dictionary, in addition to providing the solution in a well-planned way, it should also be made from scratch, not from an available printed version, and it should be updated regularly, according to recent developments and inputs from users.

5. Conclusion

The discussion in this article shows that the current Internet finance dictionaries have not utilized the Internet technological features well in designing their access routes. Some Internet finance dictionaries only use access routes similar to those in printed dictionaries, so they can simply be called meagre Internet dictionaries. Other Internet finance dictionaries utilize so many technological features that users still incur high lexicographical information costs. Such dictionaries can be called extravagant Internet dictionaries owing to the fact that they carelessly use too many technological features.

The discussion also shows that using a theoretical approach such as the modern theory of lexicographical functions results in better designs for future Internet finance dictionaries. Knowing the specific types of potential user, user situation, and user need will enable lexicographers to create dictionaries which can provide better and more focused assistance to cover the needs of the users. Therefore, modern lexicographers should always aim at creating state-of-the-art dictionaries which can be categorized as proper Internet dictionaries or real Internet dictionaries.

References

- Andersen, B. and S. Nielsen.** 2009. Ten Key Issues in Lexicography for the Future. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009. *Lexicography at a Crossroads*: 355-365. Bern: Peter Lang.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bergenholtz, H. and M. Johnsen.** 2007. Log Files Can and Should Be Prepared for a Functionalistic Approach. *Lexikos* 17: 1-20.
- Bergenholtz, H. and S. Nielsen.** 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bogaards, P.** 1999. Access Structures of Learners' Dictionaries. Herbst, T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)*: 113-130. Tübingen: Niemeyer.
- Carew, E.** 1996. *The Language of Money*. <http://www.anz.com/edna/dictionary.asp>. [Accessed 15 December 2009].
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
<http://www.financial-dictionary.thefreedictionary.com> [Accessed 15 December 2009].
<http://www.investopedia.com/dictionary/default.asp> [Accessed 15 December 2009].
- Morgenson, G. and C.R. Harvey.** 2002. *The New York Times Dictionary of Money and Investing*. <http://www.duke.edu/~charvey/Classes/wpg/glossary.htm>. [Accessed 15 December 2009].
- Nielsen S.** 1999. Mediostructures in Bilingual LSP Dictionaries. *Lexicographica, International Annual for Lexicography* 15: 90-113.
- Nielsen, S.** 2008. The Effect of Lexicographical Information Costs on Dictionary Making and Use. *Lexikos* 18: 170-189.
- Nielsen, S. and L. Mourier.** 2005. Internet Accounting Dictionaries: Present Solutions and Future Opportunities. *Hermes, Journal of Linguistics* 34: 83-116.
- Tarp, S.** 1994. Funktionen in Fachwörterbüchern. Bergenholtz, H. and B. Schaefer (Eds.). 1994. *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*: 229-246. Forum für Fachsprachen-Forschung 23. Tübingen: Gunter Narr.
- Tarp, S.** 1998. Leksikografien på egne ben. Fordelingsstrukturer og byggedele i et brugerorienteret perspektiv. *Hermes, Journal of Linguistics* 21: 121-137.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. *Lexicographica. Series Maior* 134. Tübingen: Niemeyer.

Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries*

Robert Lew, *Department of Lexicology and Lexicography, School of English,
Adam Mickiewicz University, Poznań, Poland (rlew@amu.edu.pl)*

Abstract: One finding of user studies is that information on meaning tends to be what dictionary users want most from their dictionaries. This is consistent with the traditional image of the dictionary as a repository of meanings of words, and this is also borne out in definitions of the item DICTIONARY itself as given in dictionaries. While this popular view has not changed much, the growing role of electronic dictionaries can change the lexicographers' approach to meaning representation. Traditionally, paper dictionaries have explained words with words, using either a definition or an equivalent, and occasionally a line-drawn picture. However, a prominent feature of the electronic medium is its multimodality, and this offers potential for the description of meaning. While it is much easier to include pictorial content, electronic dictionaries can also hold media objects which paper cannot carry, such as audio, animation or video. Publishers are drawn by the attraction of these new options, but are they always functionally useful for the dictionary users? In this article, the existing evidence is examined, and informed guesses are offered where evidence is not yet available.

Keywords: ELECTRONIC DICTIONARY, MEANING, ILLUSTRATION, ANIMATION, AUDIO, SOUND EFFECTS, VIDEO, MULTIMODALITY, SPECIALIZED LEXICOGRAPHY, LEARNER'S DICTIONARY

Opsomming: Multimodale leksikografie: Die voorstelling van betekenis in elektroniese woordeboeke. Een bevinding van gebruikerstudies is dat inligting oor betekenis dit blyk te wees wat woordeboekgebruikers die meeste van hul woordeboeke verlang. Dit is in ooreenstemming met die tradisionele beeld van die woordeboek as 'n bewaarplek van die betekenis van woorde, en dit word ook bevestig deur die definisies van die item WOORDEBOEK self soos aangebied in woordeboeke. Terwyl hierdie populêre siening nie veel verander het nie, kan die toenemende rol van elektroniese woordeboeke leksikograwe se benadering tot betekenisvoorstelling verander. Tradisioneel het papierwoordeboeke woorde met woorde verduidelik deur óf 'n

* This article includes some ideas and content from two plenary lectures of the author presented at the First International Symposium on Lexicography and L2 Teaching which took place at the Guangdong University of Foreign Studies, Guangzhou, China, 26–28 November 2008, and at the Fifteenth International Conference of the African Association for Lexicography which took place at the University of Botswana, Gaborone, Botswana, 19–21 July 2010.

definisie, óf 'n ekwivalent te gebruik, en soms 'n lyngetekende illustrasie. 'n Opvallende kenmerk van die elektroniese medium is sy multimodaliteit, en dit bied 'n moontlikheid vir die beskrywing van betekenis. Terwyl dit baie makliker is om illustratiewe inhoud in te sluit, kan elektroniese woordeboeke ook mediaobjekte bevat wat papier nie kan oordra nie, soos oudio, animasie en video. Uitgewers word aangetrek deur hierdie nuwe keuses, maar is hulle altyd funksioneel nuttig vir die woordeboekgebruikers? In hierdie artikel word bestaande getuienis ondersoek en ingeligte raaiskote aangebied waar getuienis nog nie beskikbaar is nie.

Slutelwoorde: ELEKTRONIESE WOORDEBOEK, BETEKENIS, ILLUSTRASIE, ANIMASIE, OUDIO, KLANKEFFEKTE, VIDEO, MULTIMODALITEIT, GESPECIALISEERDE LEKSIKOGRAFIE, AANLEERDERSWOORDEBOEK

1. Introduction

1.1 Lexicography: words and beyond

Traditionally (as well as etymologically), *lexicography* is about *words*. Words (in the various technical senses of this general-language term) are the primary objects of description of dictionaries. But words are not just the thing being described; they have also remained the most important instrument of the dictionary-maker in the very job of describing. Words, then, have figured prominently on both sides of the lexicographic equation.

The questions addressed in this article are the following: How does the current transition of dictionaries to the electronic medium affect the role of the traditional verbal orientation of lexicography? Will the shift to the electronic mode lead to words being used in a different way in lexicographic explanation? Finally, how can dictionaries transcend words and employ other modalities: static and animated graphics, audio recordings, video sequences?¹

1.2 Meaning as a central notion in lexicography

Dictionary users have varying needs when consulting dictionaries, and those needs are naturally dependent on personal preferences, but also, or perhaps primarily, on the specific task in which dictionary users happen to be involved, and circumstances of consultation. This much is clear from the accumulated body of user research, but what we also learn from such research is that there is one consultation need that clearly dominates, whatever the users' mother tongue, nationality, or proficiency level. What users want most from their dictionaries much of the time is *meaning*. This finding finds support in dictionary definitions themselves, such as the one given under the entry **dictionary** in the (online version of the) Cambridge Dictionary for Advanced Learners (CALD2, Walter 2005):

dictionary

noun [C]

1 a book that contains a list of words in alphabetical order with their meanings explained or written in another language, or a similar product for use on a computer

If we agree that meaning is a key element in a lexicographic work, it follows that lexicographers should give their most careful attention to the treatment of meaning, so that they can satisfy the primary need of dictionary users. Traditional printed dictionaries have used a repertoire of devices for representing meaning in paper dictionaries, most of them having to do with words. As electronic dictionaries grow in importance, this repertoire can be extended, but also some of the traditional modes of meaning provision may be used in somewhat different ways. I will begin with the latter issue, gradually expanding the range of explanatory modes as we move on.

2. Modes of meaning indication in dictionaries

2.1 Verbal mode

The dominant way of indicating meaning in dictionaries has been *verbal explanation*, which relies on words. Such explanation, however, can take a number of forms which can be used alone or in combination.

2.1.1 Definition

Definition is clearly the most common and prototypical carrier of meaning in monolingual dictionaries; normally, a definition is a paraphrase of a lexical unit being defined by means of a more elaborate syntactic construction, but a number of interesting variations are possible (Wierzbicka 1985; Hanks 1987; Lew and Dziemianko 2006; Fabiszewski-Jaworski and Grochocka 2010). Definition has managed to stand the test of time: as such, it has endured for many centuries of lexicographic practice across a variety of dictionary cultures, and so there is no good reason why it should not be allowed to play a prominent role in modern electronic dictionaries. Still, it is interesting to reflect on its role in the new technological context afforded by electronic dictionaries.

2.1.1.1 Defining vocabulary

In the very popular English learners' dictionaries, definitions are now usually written using a restricted defining vocabulary of some twenty five hundred (give or take a few hundred) items (for a detailed historical account of the so-called vocabulary control movement, see Cowie 1999). The chief motivation behind the idea is that restricted-vocabulary definitions should be easier to understand for foreign learners, although there may be unexpected benefits to

native speakers as well (McCreary and Amacker 2006). However, there are down sides, too, and here is one of them: by limiting the choice of words you can use in the definition to the basic set of most common words in their most common meanings, you tend to lose some of the discriminating power to tease out the finer shades of meaning. In actual lexicographic practice, there are cases when it is extremely difficult to avoid using a word from outside the defining vocabulary set. A case in point is the lemma **lava**, which can hardly be defined successfully without a direct reference to a volcano. Seasoned lexicographers are aware of the problem, of course, so in cases like this, rather than talk about some 'mountain with a hole', they will deviate from the principle and nevertheless use the forbidden *volcano*. The off-the-list status of this word, however, will often be indicated with the use of special typography, commonly lower capitals. In addition, the relevant sense may also be marked with a raised number if the capitalized term happens to have a polysemous entry. This special treatment formally indicates that the word used in the definition lies outside the defining vocabulary set, but also, on a more practical level, signals to the users that they may well find it necessary to cross-refer to the entry for the capitalized term. However, the new factor in the equation is that such an act of cross-reference is, in a well-designed electronic dictionary, a very different operation compared to its traditional paper-book precursor: what used to be laborious page-turning and letter hunting, can now become a simple click of the mouse or even as little as hovering your mouse over the target, with a small popup window providing an instant explanation. In addition, the smart e-dictionary interface will reduce the word to its lemma form, will never mistake the headword for a similar one, and will not get distracted by an irrelevant entry.

Given the relative ease with which instantaneous assistance can now be given (in principle at least, if not always in practice) for problematic words in the definition, it might make good sense to make more liberal use of words outside the strict defining set in an electronic dictionary than in a paper dictionary. Or, perhaps there is room in electronic dictionaries for larger defining vocabulary sets. Care is advised here, as too many infrequent words in a definition may create comprehension problems. We will revisit problems of dictionary definitions and how to help overcome them in section 2.2 below, but now let us move on to another salient lexicographic verbal element: the equivalent.

2.1.2 Equivalent

Just as definition is the primary instrument of meaning provision in monolingual dictionaries, so is the equivalent in bilingual dictionaries (see Adamska-Sałaciak 2006; 2010 for a detailed discussion of equivalents in bilingual dictionaries). Although the transition to the electronic medium is not likely to radically change lexicographers' way of thinking about dictionary equivalents, recent research (Laufer and Levitzky-Aviad 2006) suggests that L1→L2 electronic dictionaries can potentially benefit from adding instant access from the

target language equivalents normally given in the entry to more complete information about these items, such as details on their meaning (a similar proposal is made in Bogaards and Hannay 2004).

However, while such enhanced entries might assist in text production and language study, looking up meaning does not seem to be the most salient component of foreign language text production, a process which basically proceeds from concepts (often L1-encoded) to foreign language form. In contrast, meaning indication is primarily relevant in bilingual dictionaries going from L2 to L1, normally used in text decoding. Here, equivalents appear to remain the unchallenged leader when it comes to the job of conveying meaning (Laufer and Hill 2000; Lew and Doroszewska 2009), just as they do for paper dictionaries (Lew 2004).

The above comments do not mean that the electronic medium has nothing to offer in the way of enhancing meaning indication. One case in point would be to design new ways to guide the user to the relevant sense. This is the topic of a recent experimental study by Lew and Tokarek (2010), where the authors test three versions of an experimental online interface to a comprehensive Polish-English dictionary for polysemous entries. The three versions differed in terms of entry navigation devices used. Version One was a plain entry with no extra navigation assistance. Version Two added an entry menu at the top of each entry, as a list of senses accompanied by abbreviated cues, and hyper-linked so that clicking on the specific sense scrolled the display down to the sense proper. Version Three included an entry menu just like that in Version Two, but in addition the target sense in the body of the entry was shown as highlighted against a yellow background, once it was selected from the entry menu. This last version was found to be significantly faster, with a mean access time of 25.6 seconds, compared to 33.2 seconds for the plain menu version and 34.1 seconds for the version with no menu. In addition, the version with sense highlighting yielded mean error rates in the translation task used in the study of about half the magnitude of those in the two simpler interfaces. These findings suggest that target sense highlighting, simple as it is, is a very efficient navigation device even in bilingual dictionaries, where the presence of L1 elements in entries makes scanning long entries less of a problem than would be the case for monolingual dictionaries in a foreign or second language. What is also clear is that it is no longer enough to talk about 'paper' versus 'electronic' dictionaries, but that efficiency will largely depend on the particular solutions adopted (see for instance Tan 2009; Chen 2010; Dziemianko 2010).

2.1.3 Example

Examples in dictionaries have a whole range of functions and uses (Fox 1987; Toope 1996; Purczyńska 2002), including some that may not be immediately obvious to users themselves, such as providing guidance on grammar by way of pattern illustration (Bogaards and Van der Kloot 2001, 2002). However, within the focus of this article, the contribution of exemplification to meaning

indication is most relevant, and in this context it is worth to refer to studies which have found examples and definition to be mutually supportive (Summers 1988; Laufer 1993). Still, a recent study (Al-Ajmi 2008) finds the opposite, so the matter clearly needs further study. One factor that may be responsible for this puzzling disparity may be the quality and type of examples, and this brings us to the issue of the criteria for example selection.

In recent decades, there has been quite a heated debate on the role of corpora in providing examples. Three broad approaches can be identified: corpus-derived (examples taken out of corpora with minimal modification), corpus-based (adapted from corpus material), and invented, with arguments being presented for — and against — each of the above (Fox 1987; Cowie 1989; Laufer 1992; Humblé 1998, 2001). Currently, editors of English monolingual learners' dictionaries at least seem to have reached a sort of compromise, mostly going for corpus-based, but modified examples.

While it is generally accepted that text corpora are a good source of examples to be included in a dictionary at the compilation stage, corpora themselves can now form part of an electronic dictionary, and thus offer a useful on-line source of additional examples. Ideally, users should be granted access to additional corpus examples from all the relevant points within an entry, typically in the form of a concordance or some variation thereof. Additional corpus examples may well be a satisfactory solution to the problem of restrictions on how much exemplification a dictionary can provide, due to space constraints (Xu 2008). One of the first major dictionaries to attempt this was the second edition of COBUILD (Sinclair 1995), which added to its CD-ROM version a 5-million-word sample from the WordBank corpus. However, this innovation was not very well received, perhaps as a result of the rather poor integration between the main text of the dictionary and the corpus sample, or perhaps owing to the disappointing size of the corpus itself.

In this context, it is relevant to mention the GDEX tool (Kilgarriff et al. 2008), an implementation of an algorithm designed to rank corpus citations. Currently, the tool can be accessed in an experimental demo version as part of an automated dictionary of English collocations, available at ForBetterEnglish.com. This is a very promising recent development which could provide future dictionary users with useful corpus-derived examples on demand without flooding them with too much irrelevant material.

2.2 Audio presentation of verbal elements

If an electronic dictionary has built-in audio capability, then it is in principle possible to present the verbal components of the entry (such as the headword, definition, or example), not just in conventional spelling, but as the spoken word. In this case, *spoken* could refer to voice recordings of a human reader, or else to synthesized speech, the latter not being very popular at this point, but this might change in the near future as speech synthesis techniques get better. While it is becoming something of an industry standard to offer audio re-

cordings of the headword itself, users could conceivably benefit from audio representation of other microstructural elements, notably the definition and examples.

Apart from their unquestioned attractions, spoken definitions and examples have less immediately obvious potential benefits. We have already (section 2.1.1 above) pointed to some problems with the use of restricted defining vocabulary in learners' dictionaries. Beyond those already mentioned, there is at least one more disadvantage: simplifying the lexis often requires compensatory strategies on the part of the definer, so that the simpler vocabulary may come at the cost of increased grammatical complexity (such as the addition of modifying clauses) and longer definitions. In fact, any of these two can lead to parsing problems, with users experiencing difficulty segmenting the definition correctly into phrasal components. This difficulty is particularly likely to be felt by non-native speakers or learners, owing to their imperfect command of the language. For them, an opportunity to hear the definition being read out aloud, and not just printed in ordinary orthography, offers the helpful cues inherent in the prosodic characteristics of speech: stress, rhythm, and intonation. This additional layer is likely to provide assistance in conveying the syntactic-structural composition of the definition to the users, thus helping them to segment, and ultimately understand, the definition better. This in addition to the expected long-term benefits of being exposed to authentic pronunciation. The same rationale might apply to examples; in fact, LDOCE4 in its CD-ROM version already provides digitized audio recordings of examples.²

In addition, audio presentation would afford dictionary users an opportunity to get more exposure to so-called 'real language' — here in the sense of speech rather than writing, while also serving as a pronunciation model. If we endorse the general trend observable in English Monolingual Learners' Dictionaries to make definitions more like teachers talking to students, then presenting definitions in audio format is certainly much closer to teachers *talking* than the traditional printed definitions.

Much as the idea of spoken definitions seems theoretically attractive, user studies are needed to establish to what extent the benefits of definitions (and examples) in audio format are real and significant. Some positive evidence is, however, already available when it comes to audio recordings of lemmas. Laufer and Hill (2000) examined the options selected for meaning representation in an innovative experimental electronic dictionary interface by Israeli and Chinese (Hong Kong) learners of English at university level. The options available to users were: L1 translation, L2 definition, illustrative examples, etymology, 'extra' information (i.e. other forms of the word, phonemic transcription, register, complementation, related meanings, and other semantic and syntactic details), and spoken recordings of headwords. A comparison of the two groups of subjects revealed that the Chinese students referred to recorded pronunciations significantly more frequently than the Israeli subjects, and doing so helped them retain the meaning of the consulted items better. The authors explain that one possible reason for the enhanced value of spoken forms for

Chinese users might be that 'Chinese preference for the pronunciation option could somehow be related to the fact that Chinese dictionaries are arranged according to the phonetic radical and so Chinese lookup (sic) words in a dictionary by sound' (Laufer and Hill 2000: 70). This is an important observation underscoring the point that the specific solutions adopted by electronic dictionary makers need to take cognizance of the particularities of the language(s) being treated as well as the needs and habits of prospective users.

2.3 Non-verbal mode

Verbal representation of meaning has been the cornerstone of lexicographic practice for centuries, and although it remains just as important today, traditional lexicography is not exclusively about words: pictorial illustrations have also established a presence in paper lexicography. This, plus some other electronic-only options (not possible on printed paper) will be discussed below.

Before we move on to the discussion of such visual elements, however, let us dwell for a while longer on audio recordings, but this time recordings, not of spoken linguistic forms, but rather of non-linguistic sounds.

2.3.1 Audio recordings of non-linguistic sounds

Audio clips in an electronic dictionary need not be restricted to the duplication of lexicographic data in printed form. This potential of e-dictionaries was already noted by Dodd (1989: 91), who postulated the inclusion of audio clips for selected entries, specifically for onomatopoeic words. Ooi (1998: 112) argues that an actual recording of the sound of a bell would contribute more to the explanation of meaning than a mere verbal description, or a picture of it. Ooi's observation can be generalized to other lexical items for which the sound is an important and recognizable component of meaning. In fact, there are dictionaries that already offer this: the online *Macmillan English Dictionary* (Rundell 2007) includes the so-called *sound effects* (the term also originally used by Dodd). Among those, we find the recordings of sounds of musical instruments under their names, including the common ones such as GUITAR, PIANO, VIOLIN, RECORDER, as well the less-widely known instruments such as SITAR. We also find animal and bird calls, such as the roar of a lion under ROAR (but not under LION), or the hoot of an owl under HOOT (but not under OWL). Sounds made by humans, voluntary or otherwise, are also well represented (cf. the entries for CLAP, LAUGH, or HICCUP).

2.3.2 Pictorial illustration (static)

Although, by conventional wisdom, a picture is worth a thousand words, not all vocabulary items can be usefully, or indeed at all, illustrated. This is the reason why *pictorial dictionaries* will not pack a wordlist (nomenclature) compara-

ble to that of a general dictionary, and an *illustrated dictionary* will only choose to include pictorials for a selection of its entries (and senses). For instance, according to the information in the front matter (Summers 1987: F49), LDOCE2 aims to add illustrations for the following four categories of items:

- (a) common animals, plants, objects;
- (b) things not easily explained in words, such as shapes, complex actions;
- (c) groups of related words; and
- (d) basic meaning of words often used in a figurative or abstract way.

Stein (1991) gives a fuller and more detailed list of items for which lexicographic illustrations are used, based on her analysis of several dictionaries, including two English learners' dictionaries. Ilson (1987), on the other hand, adopts a broader concept of dictionary illustration which also includes tables and diagrams. There appears to be, then, a fairly good understanding of what types of items may potentially be complemented with illustrations. The practical question, though, is whether adding pictures offers real benefits to the dictionary user. In this connection, let us look next at two relevant studies.

Nesi (1998) investigated the effectiveness of illustrations in conveying the meaning of everyday household objects. Her target items were: *colander*, *insole*, *plunger*, *shoehorn* and *spout*; and all had clearly recognizable visual features. In her experiment, Nesi observed that her subjects, in the majority undergraduate students who were not native speakers of English, were frequently at a loss identifying actual meanings based on definitions and examples alone, but they would often correctly identify the object once confronted with a picture. This finding indicates that verbal explanation of meaning by way of definition and example can benefit from the addition of pictorial illustrations.

In her study, Nesi did not offer subjects an equivalent in their native language. This was done in a more recent study by Gumkowska (2008), which focused on vocabulary retention by twenty Polish learners of English from bilingual-dictionary-like concrete noun entries with and without pictures, again using concrete noun vocabulary items. Gumkowska (at the advice of the present author in the role of thesis supervisor) adopted a cross-balanced design, so that one version of the task included pictures for half the entries, while the other version illustrated the other entries. This way, each subject was exposed to both illustrated and unillustrated types of entries, and each entry was presented under two conditions in equal measure, thus compensating for the confounding influence of the effects of item and subject. I have calculated the retention rate for the picture-enhanced headwords to be close to 80%, compared with 70% for the entries with bare Polish translations, and this difference turned out to be statistically significant by a paired T-test ($T_{(df=38)} = 2.58$, $p = 0.014$), despite the small sample size.

The two studies point to double benefits of pictorial illustrations as an additional meaning indicator in dictionaries, at least for concrete nouns. As shown by Nesi, pictures help with immediate recognition and comprehension,

which is often the primary function of dictionaries. But the benefits do not stop there, as the presence of pictures also promotes vocabulary acquisition in the foreign language, as evidenced by Gumkowska's study.

If illustrations are indeed as effective as the two studies indicate, the natural course for lexicographers to take would be to try and include a larger number of illustrations in dictionaries. In paper dictionaries, this option has always been constrained by two factors: cost and storage space. Illustrations tend to be more expensive and problematic in the process of typesetting and printing. They also cannot be very small, or else they will lose clarity and the benefits will be diminished or lost. In a printed book, every single picture included adds to the size of the volume, and this, indirectly, again impacts cost, as well as making the dictionary bulky and unwieldy. But with the large and expanding storage capacity of most types of electronic dictionaries (standalone devices and resident dictionaries for mobile phones may be something of an exception), including pictorial illustrations for a larger number of entries and senses than has been usual in paper dictionaries is a compelling prospect. However, one should not forget that the total *storage space* of a dictionary is not the same as its *presentation space* (roughly, how much data it can display to the user at a time; see Lew (in press) for details). The latter is still about as restricted as it is in paper dictionaries, and more so on handheld devices, including mobile-phone-based dictionaries. The fundamental technical problem at present is the difficulty of combining portability and small device size with a comfortably large display. Perhaps a radical solution becomes available in the near future in the form of projector glasses, which would display an image on the inside of their lenses for the wearer to view, thus utilizing a large fraction of the human field of vision. In the not-so-near future, 3D holographic projectors may become a possibility, but for the moment this is the stuff of science *fiction* rather than science.

Another issue which may be affected by technological change is the use of colour in illustrations. Colour printing tends to be expensive (though it has become more affordable in recent years); in contrast, colour displays are now the default on electronic devices except in some types of handheld portables, and in e-book readers. This means that cost ceases to be an important consideration in choosing between black-and-white and colour illustrations, and the latter can be used more liberally. Having said that, there is no evidence that full-colour illustrations are in fact more effective in dictionaries than the more traditional simple iconic line drawings. This last point is related to another choice: that between drawings and photographs, which we now turn to.

2.3.3 Photographs

With today's easy-to-use and efficient digital cameras, photographs may often be the quicker, cheaper, and easier choice than drawings, at least for everyday objects within easy reach. Photographs can be included in electronic dictionar-

ies more readily than on paper, particularly if the device is equipped with a colour screen. However, photographs may not be without problems. In terms of prototype theory, it is probably easier to represent a 'prototypical' bird, chair or car, than it is by means of a photograph. The latter usually includes a lot of detail, and there is the real danger of the user interpreting some irrelevant detail as criterial for the concept being defined. Further, a hand drawing can emphasize what the focus is, i.e. which part of the pictorial illustration represents the object (and lexical item) in question. The artist can foreground and background the various elements of the picture as needed, providing subtle clues to the user consulting the entry. For example, to properly represent a **fin** in an illustration, it has to be shown in the context of a fish as a whole; at the same time, users need an indication that the word does not denote the whole fish. A professional artist can utilize a variety of techniques such as broken lines, shading and pointing arrows singling out components of the picture. Figure 1 below illustrates how this could be done. Still, manipulating a photograph to achieve a similar effect, such as adding an arrow pointing at the designate in the picture, is also possible, and occasionally one does come across such techniques in dictionaries.

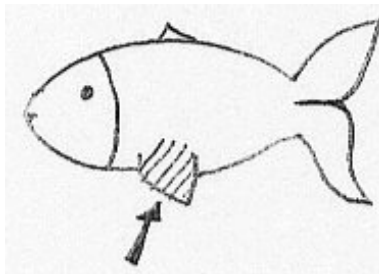


Figure 1: An illustrative picture for the entry **fin**. Note the use of shading and an arrow as foregrounding devices to prevent the mistaken identification of the lemma sign with the designate of the holonymic **fish** (after Gumkowska 2008).

2.3.4 Graphs

Iconic illustrations and photographs may be useful for illuminating the meaning of concrete objects, but what about words expressing more abstract relations, such as prepositions? These are the kinds of words whose lexicographic treatment even in the leading dictionaries still leaves one dissatisfied (Coffey 2006). Adamska-Salaciak (2008) advocates the inclusion of schematic graphs to represent the meaning of prepositions, and an example for the prepositions *above*, *over*, *under* and *below* is given in Figure 2. Of course, such a graph will not provide an exhaustive explanation of all the numerous senses and uses of the word, but will nevertheless present a schematic of the most central spatial senses.

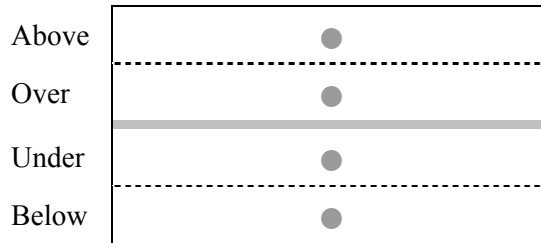


Figure 2: A graph proposed for inclusion in bilingual dictionaries as a comparison of the English prepositions *above*, *over*, *under* and *below* (Adamska-Salaciak 2008: 1482)

It would seem that graphs of this and similar types might also be useful in the representation of various meaning relations. A particularly convincing implementation of the effective visualization of meaning relations is the Visuwords interface (www.visuwords.com), here serving as a front-end to English WordNet data; but it could be used with other lexical databases as long as it can be made to 'understand' the structure. An example entry for the lemma **motor** is shown in Figure 3 below.

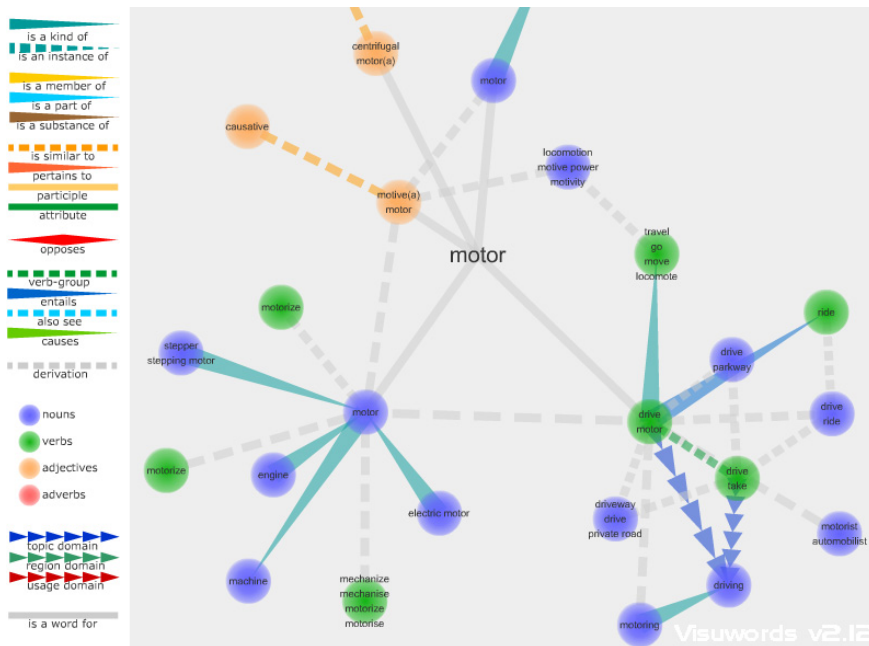


Figure 3: A floating quasi-3D graph illustrating sense relations for the entry **motor** in the Visuwords interface (<http://www.visuwords.com/?word=motor>)

2.3.5 Animation

Some of Stein's (1991: 109-111) illustratable categories actually involve, not one, but a series of static pictures. Herewith lie countable nouns shown in 'different sides or aspects ... to capture different dimensions or stages of use', an example of which is **concertina**; and perhaps even more so Stein's category 3, 'countable nouns that denote an event or an activity that is realized by a number of event or action phases', such as **eclipse** and **press-up/push up**. Another group are action verbs such as **dive**, **erupt** and **bend**. In entry types such as the above, it is hard to escape the impression that animated illustrations would form a perfect medium for conveying to the user the stages or progression of an action.

This was what Lew and Doroszevska (2009) expected to find in their study of animations in on-line dictionary entries, in which Polish learners of English were asked to read a short story online and look up unknown words in an online experimental dictionary interface. To help them understand the more difficult words, users could view any combination of the following: Polish translation, English definition, example sentences, and animated picture. Every detail of online behaviour was logged, and subjects were subsequently tested on meaning recall of the target items, which were verbs describing actions of the body and facial expressions such as **shiver**, **blush** and **frown**, believed to be well representable by means of animation. Quite unexpectedly,³ consulting animations did not turn out to have helped Polish students remember the meanings better, and in fact a strong and statistically significant *negative* effect of viewing animation on item retention was found. Why would animations have had a negative effect on retention? One possibility is that animations distracted subjects from the form–meaning relationship, which is essential for successful retention. Another possible reason is that the actual animations used in the study could have been misleading: although some initial piloting was done and no problems were found, plus all those subjects who viewed animations also consulted Polish equivalents, so it is not likely that they actually misunderstood the items. Whatever the reason, the outcome invites the recommendation that animations in future electronic dictionaries be used cautiously and sparingly, at least before more evidence becomes available on the effectiveness of animations for different types of entries.

2.3.6 Video clips

Video sequences are the most demanding type of lexicographic data in terms of storage space; plus, they tend to be expensive and difficult to produce. Also, for the user to watch a video clip takes some time; this means that the consultation act cannot be quick, which is a fundamental consideration in dictionary use, at least for Hausmann's *selective* (German *punktuell*) type of dictionary consultation (Hausmann 1977: 144). Further, as pointed out by Chun and Plass (1996), the nature of video sequences is such that they are too transient for the viewer

to build a stable mental model. This is different from static pictures, which can be examined for as long as the user wishes. The risk of a mismatch between the user's cognitive pace and the timing of the visual presentation is one disadvantage that video clips appear to share with animated pictures, and this factor could account for the poor effectiveness of animations (cf. 2.3.5 above). If video clips are to be useful in dictionaries, it is likely to be in connection with the learning function of the dictionary, to provide assistance with items such as conventionalized exchanges, pragmatic pairs, etc. (e.g. asking the way). Videos, with their rich contextualization, can convey useful information on the pragmatic and situational context of linguistic communication. Some video sequences have already been included in learners' dictionaries, such as in LDOCE4 (Summers 2003), although somehow they are no longer offered in the latest fifth edition (Mayor 2009).

3. Closing comments

One area of lexicography which could perhaps benefit significantly from the multimodal approach is LSP dictionaries. This is the approach proposed by Talib (2009) for a dictionary of narration. Talib cites filmmaking terms such as 'flashback', 'voice-over narration' and 'point-of-view shot', predicting that 'the possibility is very real that some specialist dictionaries of the future may end up not consisting of words, but videos, pictures and non-linguistic sounds, and other videos, pictures and non-linguistic sounds will be used to "define" them' (Talib 2009: 258).

Unlike Talib, I would not go so far as to expect words to surrender their leading role in lexicographic meaning explanation. However, given the technical potential of modern electronic dictionaries, various possibilities present themselves of complementing the traditional verbal explanation. The expectation is that multi-modal processing should engage the dictionary user more and bring about improved comprehension and knowledge acquisition, but in practice this is not always so. Clearly, much more research is needed before we are able to identify the optimal combination of ways of presenting meaning in dictionaries. It should not be surprising if the solutions turn out to be sensitive to factors such as dictionary culture, consultation goals and context, level of proficiency in the language and specialized domain, and type of lexical item.

Endnotes

1. No systematic attempt is made in this article to address other aspects of electronic lexicography; the reader could study De Schryver (2003) on the subject.
2. Audio recordings are also available in the free online version at <http://www.ldoceonline.com>, but only for words beginning with the letters D and S.
3. It is the subversion of such seemingly inescapable expectations that underscores the importance of experimental user research, *pace* Tarp (2009).

References

- Adamska-Sałaciak, A.** 2006. *Meaning and the Bilingual Dictionary: The Case of English and Polish*. Frankfurt am Main: Peter Lang.
- Adamska-Sałaciak, A.** 2008. Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look. Bernal, E. and J. DeCesaris (Eds.). 2008: 1477-1485.
- Adamska-Sałaciak, A.** 2010. Examining Equivalence. *International Journal of Lexicography*: Advance Access published online on July 23, 2010.
- Al-Ajmi, H.** 2008. The Effectiveness of Dictionary Examples in Decoding: The Case of Kuwaiti Learners of English. *Lexikos* 18: 15-26.
- Bernal, E. and J. DeCesaris (Eds.).** 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra.
- Bogaards, P. and M. Hannay.** 2004. Towards a New Type of Bilingual Dictionary. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004, Volume 2*: 463-474. Lorient: Université De Bretagne Sud.
- Bogaards, P. and W.A. Van der Kloot.** 2001. The Use of Grammatical Information in Learners' Dictionaries. *International Journal of Lexicography* 14(2): 97-121.
- Bogaards, P. and W.A. Van der Kloot.** 2002. Verb Constructions in Learners' Dictionaries. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002*: 747-757. Copenhagen: Center for Sprogteknologi, Copenhagen University.
- Chen, Y.** 2010. Dictionary Use and EFL Learning. A Contrastive Study of Pocket Electronic Dictionaries and Paper Dictionaries. *International Journal of Lexicography*: Advance Access published online on May 24, 2010.
- Chun, D.M. and J.L. Plass.** 1996. Effects of Multimedia Annotations on Vocabulary Acquisition. *Modern Language Journal* 80(2): 183-198.
- Coffey, S.** 2006. High-Frequency Grammatical Lexis in Advanced-Level English Learners' Dictionaries: From Language Description to Pedagogical Usefulness. *International Journal of Lexicography* 19(2): 157-173.
- Cowie, A.P.** 1989. The Language of Examples in English Learner's Dictionaries. James, G. (Ed.). 1989: 55-65.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Dodd, W.S.** 1989. Lexic computing and the Dictionary of the Future. James, G. (Ed.). 1989: 83-93.
- Dziemianko, A.** 2010. Paper or Electronic? The Role of Dictionary Form in Language Reception, Production and the Retention of Meaning and Collocations. *International Journal of Lexicography*: Advance Access published online on February 1, 2010.
- Fabiszewski-Jaworski, M. and M. Grochocka.** 2010. Folk Defining Strategies vs. Comprehension of Dictionary Definitions: An Empirical Study. Kernerman, I. and P. Bogaards (Eds.). 2010. *English Learners' Dictionaries at the DSN 2009*: 89-105. Tel Aviv: K Dictionaries.
- Fox, G.** 1987. The Case for Examples. Sinclair, J. (Ed.). 1987: 137-149.
- Gumkowska, A.** 2008. *The Role of Dictionary Illustrations in the Acquisition of Concrete Nouns by Primary School Learners and College Students of English*. Unpublished B.A. Thesis. Szczecin: Collegium Balticum.
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J. (Ed.). 1987: 116-136.

- Hausmann, F.J.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.
- Humblé, P.** 1998. The Use of Authentic, Made-up and 'Controlled' Examples in Foreign Language Dictionaries. Fontenelle, T., P. Hiligsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *EURALEX '98 Proceedings. Papers Submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 593-599. Liège: English and Dutch Departments, University of Liège.
- Humblé, P.** 2001. *Dictionaries and Language Learners*. Frankfurt am Main: Haag und Herchen.
- Iison, R.F.** 1987. Illustrations in Dictionaries. Cowie, A.P. (Ed.). 1987. *The Dictionary and the Language Learner*: 193-212. Lexicographica Series Maior 17. Tübingen: Niemeyer.
- James, G. (Ed.)**. 1989. *Lexicographers and their Works*. Exeter Linguistic Studies 14. Exeter: Exeter University Press.
- Kilgariff, A., M. Husak, K. McAdam, M. Rundell and P. Rychlý.** 2008. GDEX: Automatically Finding Good Dictionary Examples in a Corpus. Bernal, E. and J. DeCesaris (Eds.). 2008: 425-432.
- Laufer, B.** 1992. Corpus-Based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, K.H., K. Varantola, T. Salmi-Tolonen and J. Schopp (Eds.). 1992. *EURALEX '92 Proceedings. Papers Submitted to the 5th EURALEX International Congress, Tampere 1992*: 71-76. Tampere: Department of Translation Studies, University of Tampere.
- Laufer, B.** 1993. The Effect of Dictionary Definitions and Examples on the Use and Comprehension of New L2 Words. *Cahiers de Lexicologie* 63(2): 131-142.
- Laufer, B. and M. Hill.** 2000. What Lexical Information Do L2 Learners Select in a CALL Dictionary and How Does it Affect Word Retention? *Language Learning & Technology* 3(2): 58-76.
- Laufer, B. and T. Levitzky-Aviad.** 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' — a Dictionary for Production in a Foreign Language. *International Journal of Lexicography* 19(2): 135-155.
- Lew, R.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.
- Lew, R.** In press. Space Restrictions in Paper and Electronic Dictionaries and Their Implications for the Design of Production Dictionaries. Bański, P. and B. Wójtowicz (Eds.). In press. *Issues in Modern Lexicography*. Munich: Lincom Europa.
- Lew, R. and J. Doroszewska.** 2009. Electronic Dictionary Entries with Animated Pictures: Lookup Preferences and Word Retention. *International Journal of Lexicography* 22(3): 239-257.
- Lew, R. and A. Dziemianko.** 2006. A New Type of Folk-Inspired Definition in English Monolingual Learners' Dictionaries and Its Usefulness for Conveying Syntactic Information. *International Journal of Lexicography* 19(3): 225-242.
- Lew, R. and P. Tokarek.** 2010. Entry Menus in Bilingual Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2010. *Lexicography in the 21st Century: New Challenges, New Applications*: 193-202. Louvain-la-Neuve: Cahiers du CENTAL.
- Mayor, M. (Ed.)**. 2009. *Longman Dictionary of Contemporary English*. Fifth Edition. Harlow: Longman.
- McCreary, D.R. and E. Amacker.** 2006. Experimental Research on College Students' Usage of Two Dictionaries: A Comparison of the *Merriam-Webster Collegiate Dictionary* and the *Macmillan English Dictionary for Advanced Learners*. Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Atti Del XII Congresso Di Lessicografia, Torino, 6-9 Settembre 2006*: 871-885. Alessandria: Edizioni dell'Orso.

- Nesi, H.** 1998. Defining a Shoehorn: The Success of Learners' Dictionary Entries for Concrete Nouns. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 159-178. Lexicographica. Series Maior 88. Tübingen: Niemeyer.
- Ooi, V.B.Y.** 1998. *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Purczyńska, E.** 2002. *Examples in the Bilingual Dictionary: An Empirical Study*. Unpublished M.A. Thesis. Poznań: School of English, Adam Mickiewicz University.
- Rundell, M. (Ed.)**. 2007. *Macmillan English Dictionary*. (MED2.) London: Macmillan.
- Sinclair, J. (Ed.)**. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London/Glasgow: Collins.
- Sinclair, J. (Ed.)**. 1995. *Collins COBUILD English Language Dictionary*. Second Edition. (COBUILD2.) London: HarperCollins.
- Stein, G.** 1991. Illustrations in Dictionaries. *International Journal of Lexicography* 4(2): 99-127.
- Summers, D. (Ed.)**. 1987. *Longman Dictionary of Contemporary English*. Second Edition. (LDOCE2.) Harlow: Longman.
- Summers, D.** 1988. The Role of Dictionaries in Language Learning. Carter, R. and M. McCarthy (Eds.). 1988. *Vocabulary and Language Learning*: 111-125. London: Longman.
- Summers, D. (Ed.)**. 2003. *Longman Dictionary of Contemporary English*. Fourth Edition. (LDOCE4.) Harlow: Longman.
- Talib, I.S.** 2009. Towards a Multimodal Dictionary of Narration. Ooi, V.B.Y., A. Pakir, I.S. Talib and P.K.W. Tan (Eds.). 2009. *Perspectives in Lexicography: Asia and Beyond. New and Selected Papers from the Fourth Asialex International Conference, National University of Singapore. Papers on Lexicography and Dictionaries, Volume 1*: 249-258. Tel Aviv: K Dictionaries.
- Tan, K.H.** 2009. How Effective Is the Electronic Dictionary in Sense Discrimination? *Lexikos* 19: 262-274.
- Tarp, S.** 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.
- Toope, M.** 1996. *Examples in the Bilingual Dictionary*. Unpublished M.A. Thesis. Ottawa: University of Ottawa.
- Walter, E. (Ed.)**. 2005. *Cambridge Advanced Learners' Dictionary*. Second Edition. (CALD2.) Cambridge: Cambridge University Press.
- Wierzbicka, A.** 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.
- Xu, H.** 2008. Exemplification Policy in English Learners' Dictionaries. *International Journal of Lexicography* 21(4): 395-417.

The *Tshivenda*–English *Thalusamaipfi* / Dictionary as a Product of South African Lexicographic Processes

Mbulungeni Madiba, *Multilingualism Education Project, Centre for Higher Education Development, University of Cape Town, Cape Town, South Africa*
(mbulungeni.madiba@uct.ac.za)

and

Dion Nkomo, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch and Multilingualism Education Project, Centre for Higher Education Development, University of Cape Town, Cape Town, South Africa* (deeouf@yahoo.co.uk)

Abstract: The publication of a dictionary is regarded as the result of a lexicographic process. Three subtypes of a lexicographic process have been noted, namely the primary comprehensive, the secondary comprehensive and the dictionary specific lexicographic processes. In South Africa, the three lexicography processes correspond to the respective mandates of the Pan South African Language Board (PanSALB), the National Lexicography Units (NLUs) and the editorial teams involved in the compilation of the specific dictionaries. This hierarchical arrangement of the lexicographic practice is supported by the government within the country's national multilingual policy which was lauded in linguistic and lexicographic circles as a triumph for cultural democracy. It is almost a decade since these planned lexicographic processes have been in place. It seems the right time to consider the products of these South African lexicographic processes which are envied by many foreign lexicographers, especially in Africa. Accordingly, the article evaluates these lexicographic processes with special reference to the *Tshivenda*–English *Thalusamaipfi* / Dictionary. Specifically, it addresses the question: To what extent does this dictionary represent lexicographic development in the indigenous South African languages which were marginalised before the establishment of the NLUs? A few insights are drawn from modern lexicographic theories for the general improvement of future lexicographic practice in languages with limited lexicographic tools such as Venda.

Keywords: LEXICOGRAPHIC PROCESS, LEXICOGRAPHIC PLANNING, PANSALB, NATIONAL LEXICOGRAPHY UNITS, LEXICOGRAPHIC PRACTICE, METALEXICOGRAPHY, DICTIONARY, BILINGUAL DICTIONARY, MACROSTRUCTURE, MICROSTRUCTURE

Opsomming: Die *Tshivenda*–English *Thalusamaipfi* / Dictionary as 'n produk van Suid-Afrikaanse leksikografiese prosesse. Die publikasie van 'n woordeboek word beskou as die resultaat van 'n leksikografiese proses. Drie ondersoorte van 'n leksikografiese proses is onderskei, naamlik die primêr omvattende, die sekondêr omvattende en die

woordeboek-spesifieke leksikografiese prosesse. In Suid-Afrika stem die drie leksikografiese prosesse ooreen met die onderskeie mandate van die Pan-Suid-Afrikaanse Taalraad (PanSAT), die Nasionale Leksikografie-eenhede (NLE's) en die redaksiespanne betrokke by die samestelling van die bepaalde woordeboeke. Hierdie hiërargiese rangskikking van die leksikografiese praktyk word ondersteun deur die regering binne die land se nasionale veeltalige beleid wat in taalkundige en leksikografiese kringe geloof is as 'n triomf vir kulturele demokrasie. Dit is byna 'n dekade sedert hierdie beplande leksikografiese prosesse in plek is. Dit lyk na die regte tyd om die produkte van hierdie Suid-Afrikaanse leksikografiese prosesse te beskou wat deur baie vreemde leksikograwe, veral in Afrika, beny word. Gevolglik beoordeel die artikel hierdie leksikografiese prosesse met spesiale verwysing na die *Tshivenda-English Thalusaṁaṁṁfi/Dictionary*. Dit roer veral die vraag aan: Tot watter mate verteenwoordig hierdie woordeboek leksikografiese ontwikkeling in die inheemse Suid-Afrikaanse tale wat gemarginaliseer is voor die totstandkoming van die NLE's? 'n Aantal insigte word verkry van moderne leksikografiese teorieë vir die algemene verbetering van die toekomstige leksikografiese praktyk in tale met beperkte leksikografiese gereedskap, soos Venda.

Slutelwoorde: LEKSIKOGRAFIESE PROSES, LEKSIKOGRAFIESE BEPLANNING, PAN-SAT, NASIONALE LEKSIKOGRAFIE-EENHEDE, LEKSIKOGRAFIESE PRAKTYK, METALEKSIKOGRAFIE, WOORDEBOEK, TWEETALIGE WOORDEBOEK, MAKROSTRUKTUUR, MIKROSTRUKTUUR

1. Introduction

Gouws and Prinsloo (2005: 9) write:

The publication of any dictionary should not only be the result of the preceding compilation activities but it has to be regarded as the culmination of a much more comprehensive set of activities, the so-called lexicographic process. The compilation and eventual publication of any dictionary form part of at least one lexicographic process.

A lexicographic process is "part of a comprehensive historical process which coincides with the development of a language" (Gouws 2001: 65). It is constituted by all the activities leading to the publication of a dictionary as a text (Gouws and Prinsloo 2005: 9). Within the general theory of lexicography (Wiegand 1984: 15), the concept of a lexicographic process may be located within the second constituent theory, namely the theory of organisation. This pertains to lexicographic planning. Planning has been regarded as an important but quite often neglected element of lexicographic practice (Alberts 1999; Gouws 2001, 2003; and Gouws and Prinsloo 2005).

In South Africa, lexicographic planning occurs at both macro and micro level. At macro level, lexicographic planning is done by government through the Pan South African Language Board (PanSALB). Lexicographic planning at micro level is done by the National Lexicography Units (NLUs), the Editors-in-chief and their respective editorial teams. At this micro level, planning provides the lexicographer with an opportunity to preview the lexicographic prac-

tice and prepare for the tasks ahead and their characteristic challenges. It also contributes to the user-friendliness of dictionaries. Dictionaries produced within a well-conceived lexicographic process are characterised by predictability, calculability, analysability and controllability (Gouws 2001: 64). This means that the concept of a lexicographic process provides guidance to the lexicographer, the dictionary user and the dictionary critic in an equally useful way. It is such an essential element of lexicography that when a dictionary fails to be the effective tool that it ought to be, it is quite often regarded as the result of an ill-conceived lexicographic process, or an absolute absence of planning (Gouws and Prinsloo 2005: 9).

Metalexicography has identified the primary, the secondary and the dictionary-specific lexicographic processes as the three subtypes of a lexicographic process (Gouws 2001, 2003; and Gouws and Prinsloo 2005). Since these have been comprehensively discussed in the cited works, this article mainly focuses on the *Tshivenda–English Ṭhalusamaip̣fi/Dictiona*ry (henceforth TETD) as a product of such processes in South Africa. However, it is not only inevitable but also necessary for background information to discuss the agents of lexicographic processes in the country. These agents are PanSALB, the NLUs, lexicographers and dictionary publishers. Information on the agents of lexicographic processes in South Africa is already available in other published works (cf. Gouws 1996, 2001, 2003; Gouws and Prinsloo 2005; Kumalo 1999; Madiba 2002; and Mongwe 2006). However, in the available literature, this has been done retrospectively to or immediately after the establishment of the NLUs when the authors were generally in an optimistic mood. The exceptions were notably Madiba (2002) who raises critical questions regarding the government's involvement in lexicography and Gouws (2003) who considers a number of potentially negative factors. Overall, what remains missing is an introspective and qualitative evaluation of the operations and products of the South African lexicographic processes close to a decade since their inception. Therefore, the analysis of the TETD in this article, as well as reference to pre-NLUs and other dictionaries, gives another angle to the entire picture.

2. Agents of lexicographic processes in South Africa

South Africa is among the very few countries in the world where lexicography has been officially recognised as a professional enterprise with a potential to advance national goals. This occurred in the aftermath of apartheid and the formulation of a multilingual national language policy which recognised nine indigenous African languages as official languages, in addition to English and Afrikaans. Lexicography was rightfully identified as one important way of developing the formerly marginalised languages towards the implementation of the national language policy (Gouws 2003, Madiba 2002). Through PanSALB, which was given a mandate for establishing NLUs for each official language, the South African government facilitated the establishment of the coun-

try's lexicographic processes. PanSALB was established as an independent statutory body by an Act of Parliament (Act 59 of 1995) with the following explicit aims:

- (1) to promote respect for and ensure the implementation of the following principles:
 - (a) the creation of conditions for the development and for the promotion of the equal use and enjoyment of all the official South African languages;
 - (b) the extension of those rights relating to language and the status of languages which before 27 April 1994 were restricted to certain regions;
 - (c) the prevention of the use of any language for the purposes of exploitation, domination or division;
 - (d) the promotion of
 - (i) multilingualism; and
 - (ii) the provision of translation and interpreting facilities;
 - (e) the fostering of respect for languages spoken in the Republic other than the official languages, and the encouragement of their use in appropriate circumstances; and
 - (f) the non-diminution of rights relating to language and the status of languages existing before 27 April 1994;
- (2) to further the development of the official South African languages;
- (3) to promote respect for and the development of other languages used by communities in South Africa, and languages used for religious purposes;
- (4) to promote knowledge of and respect for the provisions and principles of the Constitution relating directly or indirectly to language matters;
- (5) to promote respect for multilingualism in general; and
- (6) to promote the utilisation of South Africa's language resources.

PanSALB had to facilitate all this through the creation of provincial and national structures which would advise on the respective official languages and activities that had to be undertaken. The subcommittee for Lexicography and Terminology played an important role in stressing that dictionaries would figure imperatively in the standardisation process (Gouws 2003: 220). Subsequently, the deliberations regarding the formation of the NLUs for each official language ensued. The NLUs were eventually established according to the PanSALB Act as amended in 1999 (Kumalo 1999, Gouws 2003). This made PanSALB an agent of the primary comprehensive lexicographic process in South Africa. PanSALB's mandate for establishing the NLUs implied several responsibilities which would have direct implications for the production of dictionaries. Since PanSALB identified "the compilation of a comprehensive monolingual explanatory dictionary" as the eventual line function of each NLU (Gouws 2003: 220), it was also its task to assist the NLUs with comprehensive planning to facilitate the achievement of that function. Lexicographically

speaking, the role of PanSALB was aptly summarised by Gouws (2003: 221) when he wrote:

The task that awaits PanSALB is to define and describe the lexicographic process to be used in South Africa.

PanSALB should realize the compelling need to devise a comprehensive plan for the South African lexicographic process.

The task of defining the primary comprehensive lexicographic process as articulated in the above quotations transcends the administrative and managerial responsibilities of the NLUs. However, Kumalo (1999: 211) wrote at the time: "PanSALB shall not impose itself on the units, but shall make it possible for them to take responsible decisions relevant to their specific and individual needs." It would appear that the formulation of the NLUs, particularly the subsuming of the objectives of the National Lexicography Units Bill and the establishment of the NLUs under PanSALB raised fears of conflicts of interests and control. While these issues are relevant constituents of the theory of organisation, which is part of the general theory of lexicography (Wiegand 1984: 15), they seem to have attracted more attention at the expense of the primary goals for the establishment of the NLUs, i.e. empowering the multilingual nation with relevant, functional and user-friendly dictionaries.

As a law, the PanSALB Act clearly outlined the objectives of each NLU:

The objectives of a unit shall be to initiate, maintain, continue, complete and from time to time improve the compilation of the dictionary and other products by:

- (a) the continuous and comprehensive collecting, arranging and sorting, in a lexicographically workable form, of the general vocabulary of the language concerned;
- (b) the editing, adaptation, and publication of the collected material according to lexicographic principles in printed and electronic form; and
- (c) the granting access to the language material and sources of the unit to researchers according to the policy of the board.

The NLUs were to operate as Section 21 Companies, with each of them managed by a National Lexicography Unit Board. The Constitution of the board was also clearly defined, making provision for:

- (a) a Chairperson appointed by the Minister;
- (b) an Editor-in-chief [...];
- (c) a person appointed by the Minister who, at the request of the Minister, has been nominated by a language body for the language concerned referred to in [...] the Pan South African Language Board Act, 1995 (Act No. 59 of 1995), to represent such body on the board;

- (d) two persons appointed by the Minister, in collaboration with the African Association for Lexicography (AFRILEX), on account of their interest in and knowledge of lexicography;
- (e) two persons appointed by the Minister in consultation with stakeholders on account of their knowledge of the language concerned as linguists or mother tongue speakers; and
- (f) a person appointed by the Minister on account of his or her marketing skills.

The functions and duties of the boards were listed in Chapter 2 of the Act, Section (6). According to Subsection (1), a board shall in addition to its other functions in terms of this Act:

- (a) formulate the policy to be followed to achieve the objectives of its unit;
- (b) govern and advise its unit in accordance with the resources at its disposal;
- (c) decide from time to time about matters relating to the publishing, printing and reprinting of the dictionary and products of its unit, including the determination of the selling price and conditions of sale of products and services of the unit; and
- (d) determine from time to time the number of review, gift, working and other copies of products of its unit to be made available free of charge.

From the foregoing, it is clear that through PanSALB, the government of South Africa has attempted to play not merely a supportive but actually a directive role as an agent of the country's primary comprehensive lexicographic process. Yet it would appear that the process has not been comprehensive enough because no further and adequate elaboration on the production of the relevant lexicographic products has been offered. Perhaps the closest is the document Regulations for the NLUs, which is equally found short of "solutions regarding problems of dictionary planning and compilation" (Gouws 2003: 227). All that could be established was that the NLUs and their boards became agents of the secondary lexicographic processes and through their editorial staff, agents of dictionary-specific lexicographic processes. In all this, metalexigraphically relevant pieces of information are missing which would be needed for a model within which the established NLUs could operate. It has to be recalled that prior to the establishment of the NLUs, some languages such as Ndebele had no lexicographic history, experience or expertise to draw from. As part of the primary comprehensive lexicographic process, prospective lexicographers had to be trained from among the linguists of the respective languages. Gouws (2003: 228) indicates that AFRILEX played an important role in this regard, but Sue Atkins, after having offered training together with Michael Rundell at SALEX 98, raises questions in a report (which Gouws substantially quotes) of whether such efforts sufficiently equipped the NLUs for general and language-specific lexicographic challenges.

The South African lexicographic processes also needed to win the confidence of the publishing industry. As indicated by Gouws (2001), the agents of the comprehensive secondary lexicographic process and the dictionary-specific lexicographic process should liaise with publishers regarding the publication of dictionaries. This was recognised from the outset when the Nasionale Boekhandel Group, the Oxford University Press (OUP) and the Southern Book Publishers were invited to the third National Language Services workshop which sought to demonstrate the viability of the South African lexicography industry as Beukes (1996) says:

If properly planned and positioned, the lexicography industry in South Africa could — as is the case elsewhere in the world — generate handsome financial benefits which could in turn play a significant role in the process of elaborating the African languages.

Representing OUP, McCallum (1996: 123) dismissed the aim of the seminar as "a somewhat narrow if not misleading view of lexicography" and argued that while there was a genuine need for lexicographic practice in the African languages, certain dictionary types would be non-viable. It is perhaps in this respect that traditional and new dictionary publishers now exist as competitors rather than collaborators with the NLU's. As far as languages are concerned, English and Afrikaans have continued to benefit, with a few commercially viable dictionaries in the African languages being compiled by freelance lexicographers. While this may seem regrettable, in the end it does not matter who produces or publishes what, as long as users are provided with products that may be efficiently used to solve the problems they face in their specific situations of use.

3. The *Tshivenda–English Ṫhalusamaipfi / Dictionary*

The Tshivenda National Lexicography Unit (TNLU) is one of the six NLU's that were established in post-apartheid South Africa (Mongwe 2006: 11), while others were simply reconstituted and named accordingly. Just like the other NLU's with regard to the languages they work on, the function of the TNLU is the compilation of Venda dictionaries. So far, the TNLU has produced only one dictionary, namely the TETD. The TETD was published in 2006, five years after the registration of the TNLU as a non-profit Section 21 Company in 2001 (Mongwe 2006). Although the writing of this dictionary was supported by PanSALB, the TNLU should take all credit and responsibility for its quality. Otherwise it would be very unfortunate if critical decisions regarding the contents and design features were imposed on lexicographers by the stakeholders whose role should be more managerial and logistic than practically lexicographic. In this regard, albeit along the line function that was determined by PanSALB for all the NLU's, the TNLU should have taken into account the fol-

lowing in their formulation of the dictionary concept and compilation of the TETD:

- the available dictionary types in the language,
- the unavailable dictionary types in the language,
- the potential target users of various types of prospective dictionaries in the language,
- the lexicographic needs of the potential users of the prospective dictionaries,
- the prioritisation, but not total neglect, of certain users and needs over others, and
- the reference skills of the users of the prospective dictionaries.

It is within this set of rubrics that the TETD is evaluated in the following subsections. Attention is given to its typology, its data categories and its structure in view of the identified target users and their needs.

3.1 A typological perspective on the TETD

The TETD is presumably a bidirectional bilingual dictionary with two macrostructural lists. The first list coordinates Venda lemmata with their English equivalents while the second list coordinates English lemmata with Venda equivalents. In this perspective, it appears better than Van Warmelo (1937) and Van Warmelo (1989), whose unidirectionality makes them more useful only to Venda speakers trying to learn English. It is outlined on the blurb that the TETD "has been compiled to meet the needs of Tshivenda Home Language learners, First Additional Language learners, Tshivenda students as well as speakers of other languages". The two macrostructures, which are arranged alphabetically, make the dictionary bi-/poly-accessible so that users have options of starting their search path using either Venda or English macrostructural entries. Prior to the TETD, one Venda dictionary having this advantage is Wentzel and Muloiwa's (1976) *Trilingual Elementary Dictionary* which has three macrostructures. In this way, the TETD clearly embodies the multilingualism agenda of South African lexicography and the post-apartheid national language policy. The availability of blurb texts in both Venda and English is consistent with this idea.

The identification of Venda home language learners, first additional language learners, Venda students as well as speakers of other languages as target users suggests that the TETD was conceived as a learner's dictionary. This observation is problematic, especially when the data categories provided in the dictionary are considered (see 3.2). The limited lexicographic treatment of lemmata can barely support language learning and other functions such as

translation. In a similar way, the blurb also describes the TETD as an explanatory dictionary. What exactly this means is difficult to understand. According to Hartmann and James (1998: 55), an explanatory dictionary "gives detailed explanations of the meanings covered". As it will be shown in the next section, this dictionary provides very limited treatment of lemmata in either macrostructures which are themselves equally limited. In the comment on the meaning slot, only equivalents and occasional brief paraphrases are provided so that they barely provide the dictionary user with adequate explanatory assistance regarding the lemmata. More details on this will be provided in the next section, on the basis of which it will be determined whether the dictionary may fulfil the functions which it purports to serve, and ultimately whether it marks a major contribution to the development of Venda lexicography in which a small number of dictionaries were already available.

Another aspect which is often used as a typological feature of dictionaries is size, resulting in typological distinctions such as pocket or pocket-size dictionaries, medium-size dictionaries, desk and multi-volume dictionaries. In modern lexicography, size probably remains an important feature only in as far as it is a function of the cost of dictionary production, cost price of the dictionary and convenience of being used in certain situations. Besides these considerations, dictionary size would be superseded by the functional value of the dictionary, which is determined by the availability and accessibility of data categories from which relevant information may be retrieved. However, it may be possible to correlate the size of a dictionary with its functional value within the parameters of a specific type of a dictionary. The two macrostructures of TETD together with a non-integrated (Gouws 2002, 2004) middle-matter add up to only 172 numbered pages. Slightly more than half of this dictionary space is allocated to the Tshivenda–English macrostructure while the remainder is taken up by the English–Tshivenda macrostructure. Even where the smaller is favoured for portability purposes, the functional value of the TETD may easily be put in doubt when compared to some of its predecessors with the same functions. Does the TETD provide more comprehensive assistance compared to Van Warmelo (1989) within its size constraints? A more thorough analysis of the TETD would confirm that it bears little, if any, significant improvements in comparison with some of the available dictionaries in Venda.

3.2 Data and information categories

An evaluation of data categories and information which may be retrieved from them is the best way of appreciating the functional value of a dictionary. Within the theory of lexicographic functions, attention is given to the relations existing between specific groups of users, the problems they encounter in certain situations, their information needs for solving the respective problems, and the kinds of information that may be retrieved from the data types available in a dictionary (Bergenholz and Tarp 1995, 2003; and Tarp 2008). It is, therefore,

not enough for the lexicographer or the dictionary publisher for that matter, to end with identifying the target users and functions of his/her dictionary in the introductory or cover entries. Relevant data categories have to be included in the dictionary so that the information which satisfies certain user needs may be retrieved. When data categories are included without this consideration, the dictionary is likely to become an object which represent less judicious copying from its predecessors, together with their failures (Landau 2001: 23). In cases where the dictionary is the first in the language, as is the case with the IsiNdebele NLU, the dictionary becomes an easily forgettable, if at all recognisable object in a linguistic community. Unfortunately, the potential users of the dictionary normally takes all blame for their lack of dictionary culture, yet such a dictionary culture may only be cultivated by the availability of dictionaries which solve the users' problems with reasonable ease. In this case, the TETD had its forerunners. The evaluation of its data types will inevitably result in its comparison with other Venda dictionaries. The focus will be on lemmata, grammar (type of speech labels) and equivalents, the only consistently provided data types in this dictionary.

3.2.1 Lemmata

Although every dictionary contains lemmata, this type of data should not be taken for granted. The lemmatised forms provide spelling information which is useful for literary text production and text reception. For the dictionary to provide optimal support with regard to these functions, important questions will concern the representativeness of the lemma entries. Firstly, was a corpus used as a dictionary basis for lemma selection? Modern lexicography is either corpus-based or corpus-aided, proper use of corpora usually resulting in representative dictionaries which reflect language as it is used. Secondly, if a corpus was used, was the frequency criterion or the predictability criterion adopted for lemma selection? The frequency criterion ensures that dictionaries capture the most frequent words, but for some users, e.g. adult native speakers, the most frequent words may not be the most sought. The use of the predictability criterion, especially in African languages, will save dictionary space by avoiding predictable inflections and derivations. However, the ability to predict will depend on whether the user is a native speaker with a good command of grammar or a second language learner with a limited grammatical competence. Thirdly, was lemma selection guided by policies which ensure that the lexical structure of a language is captured by avoiding biases towards certain word classes (such as nouns) at the expense of other categories? Ultimately, is the lemma stock included in the dictionary appropriate for the users of the dictionary and its functions? In short, lemma selection needs to be carried out in a very meticulous way so that the target users will find the words they look up as punctually as possible.

With an average of 60 lemmata per page, spread over 89 pages of the Tshi-

venḁa–English macrostructure, and considering the several less than half-printed pages at the end of certain alphabetical stretches, the TETD's Tshivenda vocabulary coverage is $\pm 5\,000$ items. This may be suitable for additional language learners, Venda students and speakers of other languages learning Venda, provided it satisfies the other questions of balance raised in the previous paragraph. It is commendable that the lemmata represent different word categories as well, although there seems to be a prevalence of nouns and verbs, which is characteristic of the earliest dictionaries of which the lemma selection was based on traditional methods. Another problem is that no information is provided on whether a corpus formed the basis for the dictionary or not. For example, if the frequency criterion, based on a corpus was used, indicating frequency would guide non-native Venda learners on the vocabulary they need to learn for basic communication purposes. This is done in the *Oxford Bilingual School Dictionary: Northern Sotho and English* (De Schryver 2008). Furthermore, a comparison of the English wordlists in the TETD and other bilingual dictionaries compiled by other NLU's working on languages with more or less similar levels of lexicographic development indicates that the hub and spoke model suggested by Gouws (2003) for bilingual dictionaries was not employed. For example, only 23 lemmata are common to both the first 60 lemma stretches of the TETD and *Dikixinari ya Xitsonga/English Dictionary*. Thus, as the editors also do not provide it, it is difficult to determine the dictionary basis of the TETD. The provision of such information would indicate the appropriateness of the dictionary for the four different types of identified target users. However, there seems to be a striking pattern when the Tshivenda–English section is compared with Murphy's (1997) online Venda dictionary. When the letter Aa alone is compared with the online dictionary, it is found that, with the exception of function words and loan words, all the lemmata in Murphy's dictionary are entered and treated similarly in the TETD. It would not be wrong for the TETD to have Murphy's and other existing dictionaries as its dictionary basis. However, what is remarkable is the minimum effort that went into making the TETD better than its predecessors, in line with the needs of the users the editors identified.

Given that the TETD was compiled at a time when several dictionaries existed in Venda, its vocabulary coverage may be far too limited for home language learners, who may be assumed to be the primary users of the dictionary. Of course, the TETD would have an advantage of containing contemporary vocabulary, but this would suggest that certain words in the older dictionaries would be excluded. Yet some such lemmata may be of cultural relevance to most target users. This brings back the issue of dictionary size, and the fact that the dictionary should have clearly prioritised certain users, either home language learners or additional language learners, because it cannot equally satisfy the needs of these totally different users. A dictionary for all is a dictionary for none or, worse still, no dictionary at all, if the functional value of the dictionary does not take precedence.

3.2.2 Grammar

The main basis for the rejection of the suggestion that the TETD may be regarded as a learner's dictionary is that, in addition to its limited macrostructure which makes it difficult for the dictionary to support vocabulary learning, very limited grammatical information is provided as part of the microstructural treatment. Only type of speech information is given. The table below presents the most used type of speech labels in the dictionary.

| Word class (English) | English Abbreviation | Word class (Venda) | Venda Abbreviation |
|----------------------|----------------------|---------------------|--------------------|
| Noun | <i>N</i> | <i>dzina</i> | <i>dzin</i> |
| Verb | <i>V</i> | <i>liiti</i> | <i>lii</i> |
| Adjective | <i>Adj</i> | <i>liṭaluli</i> | <i>liṭalu</i> |
| Preposition | <i>Prep</i> | <i>liṵhofhi</i> | <i>liṵho</i> |
| Adverb | <i>Adv</i> | <i>liḍadzisi</i> | <i>liḍa</i> |
| Conjunction | <i>Conj</i> | <i>liṭanganyi</i> | <i>liṭang</i> |
| Possessive | <i>Poss</i> | <i>liṣumbavhuṅe</i> | <i>liṵhuṅ</i> |

Type of speech information in the TETD

Type of speech labels represent important grammatical data provided in dictionaries. In the TETD, type of speech labels serve to distinguish between homographs, as is the case with the two lemmata represented by the form **anga**. After the first lemma **anga**, *lii* informs the user that the lexical item is a verb, while *liṵhu* indicates that the lexical item treated in the second lemma **anga** is a possessive. Unfortunately, some of the grammatical markers used in the dictionary articles are inconsistent with those supplied for guidance in the front matter. For example, *liṵhu* is used to indicate that **anga** may be used as a possessive equivalent to the English words *mine* or *my*, yet *liṵhuṅ* is given in the front matter as the abbreviation indicating possessives. Furthermore, grammatical markers are the only type of grammatical information the user finds in the dictionary. How the treated words combine with others in speech or writing is not provided, especially for the help of additional language learners. Example sentences, possibly derived from real texts, would have been useful. It would have been better if a grammatical section was provided to furnish non-native speakers with grammatical rules and guidance which may help them when learning to speak or write Venda. Moreover, a combined explanation giving the full form of a Venda symbol or abbreviation, together with its English equivalents with which non-native Venda learners may be familiar, as attempted in the table above, would have been useful. Otherwise the user is left struggling with the coordination of the symbols and abbreviations and their full forms in the two languages. This is not only cumbersome but also discouraging for the user.

3.2.3 Equivalentents

In both sections of the TETD, equivalentents are the main data categories provided in the comment on meaning slot. Al-Kasimi (1977: 60) distinguishes two types of equivalentents that are normally supplied by bilingual dictionaries, namely translational equivalentents and explanatory equivalentents. A translational equivalentent is "a lexical unit which can be immediately inserted into a sentence in the target language". An explanatory equivalentent is one "which cannot always be inserted into a sentence in the target language" and it tends to approximate a translational unit. Owing to linguistic and cultural anisomorphism between English and African languages such as Venda, both translational and explanatory equivalentents are provided in the TETD, the former for lexical items which signify universal phenomena and the latter in the case of culture-specific or context-dependent words.

In addition to translational and explanatory equivalentents, the TETD also treats the meaning of some lemmata by providing explanations in the target language. These explanations are brief paraphrases or descriptions which seem to be used in cases where there are no translational equivalentents. They are used either to clarify the explanatory equivalentent or on their own in the comment on meaning. A close look at the TETD shows that, although explanations are used in both sections of the dictionary, they are more prevalent in the Tshivenda–English section. The following are illustrations of the use of explanations as data from which meaning can be retrieved in the dictionary:

- aini** ... iron, an instrument used to make clothes smooth
- dzengaila** ... restless, always on the go
- founela** ... phone, make a phone call
- davha** ... work party held by one who wants to have the land ploughed or cultivated
- mbongo** ... food prepared from freshly harvested maize
- shula** ... smear the floor with cow dung

In these examples, the TETD diverts from the dominant procedure of providing meaning by exclusively using equivalentents. Instead, brief explanations which may be regarded as definitions are supplied. In the first three examples, *iron*, *restless* and *phone* are provided as equivalentents for **aini**, **dzengaila** and **davha** respectively. However, the equivalentents are supplemented by brief explanations providing the meaning in more detail and accuracy. In the case of the first and third examples, the explanations serve as meaning discriminators, because *iron* and *phone* have other meanings as well, with the former also referring to metal in general, while the latter may also be used as a noun.

With the fourth and fifth examples, the use of explanations to provide meaning is quite different from the first three examples. In the latter, the explanations are the sole data categories providing meaning. In such cases, the explanations are provided because neither translational nor explanatory equiv-

alents exist for the particular lexical items. The Venda lemmata refer to culture-specific phenomena which are probably not known in English and therefore have no English equivalents. Thus the adoption of brief explanations becomes the best strategy for presenting meaning. Although the provided paraphrases of meaning may not be easily inserted as translational equivalents for the lemmata, the explanations are given in informative sentences, leaving the user with an unequivocal idea of the meaning of lemmata. The problem, as stated earlier, is that they are used sparingly and without a clear and consistent policy. The TETD users, especially non-native learners of Venda, would actually wish that these explanations were provided in cases where lemmata are polysemous or where at least two partly synonymous equivalents are provided for lemmata. They would be an effective meaning discrimination strategy in such problematics as described in the next paragraph.

Meaning discrimination is probably one of the most difficult tasks non-native Venda learners have to contend with while using the TETD. The importance of meaning discrimination in bilingual dictionaries has been supported by scholars such as Al-Kasimi (1977), Mafela (2005) and Yong and Peng (2007). Al-Kasimi (1977: 67) notes that the need of meaning discrimination in bilingual dictionaries arises when the user is "confronted with several words which he cannot distinguish one from another". This is usually obtained owing to polysemy and synonymy. Therefore, as Yong and Peng (2007: 143) state, meaning discrimination "helps to answer the question of which sense is to be taken in the specific target language situation and guide the dictionary user towards the right or appropriate target language equivalent". Because of the absence of such help, Mafela (2005) criticises the earlier Venda dictionaries preceding the TETD. Unfortunately, the TETD, which was published a year later, is also lacking in this respect. The following examples illustrate how the TETD would have been more user-friendly had effective meaning discrimination strategies been employed.

pfa ... hear, feel, taste, understand; spit
vhafuwi ... farmers; Chief
tama ... wish, admire, desire, be eager, envy, prefer, crave

Confronted with the above articles from the TETD, a non-native Venda learner will find it difficult to choose the correct equivalent. In the first example, the lemmata **pfa** and **vhafuwi** are polysemous and their English equivalents can scarcely be used as synonyms. In the third example, the equivalents of the lemma **tama** may be regarded as synonyms. However, they cannot be used interchangeably in all contexts. All this indicates that the problems pointed out by Mafela (2005) regarding Venda dictionaries still prevail in the TETD. This suggests that in the compilation of the dictionary, adequate regard has not been given to metalexigraphy and therefore the limitations identified in the preceding Venda dictionaries have been repeated.

3.3 The TETD as a reflection of the South African lexicographic process

The TETD is a product of the lexicographic process through which a government-supported initiative of developing the country's official languages saw the establishment of NLUs for each language. For African languages such as Ndebele and Tsonga, such an initiative would see the production of the very first dictionaries. Yet in other languages such as Afrikaans, English, Xhosa and Zulu, lexicographic work has been going on for quite a long time. The establishment of formal structures for lexicographic purposes has improved the execution of lexicographic works in these languages.

While no structures existed for Venda lexicography before the establishment of the TNLU, there were at least a few lexicographic products on which to rely. Details about the forerunners of the TETD are outlined in Mawela (1999), Mongwe (2006) and Mafela (2008). These include Marole (1932, 1955, 1955a), Marole and De Gama (1936, 1954) and Van Warmelo (1937, 1958, 1989). Very little literature on these dictionaries is available. For example, none of these dictionaries has any recognisable keyness in De Schryver (2009), in which no names of Venda lexicographers are mentioned while 'Venda' and 'Tshivenḁa' carry low keyness values. According to De Schryver (2009: 389), Venda (Tshivenḁa) together with Tsonga (Xitsonga), compared to the other official South African languages, is not often discussed in lexicographic settings. It is only in *Lexikos* 9, 15 and 18 where Venda lexicographic works are discussed. Of the three articles, only Mafela (2005) discusses Venda lexicography in such a way that the quality of future dictionaries could be improved. Unfortunately, it appears to have been too late for Mafela (2005) to contribute to the quality of the TETD. The analysis of the TETD indicates that it does not reflect recent theoretical and methodological advances in lexicography. It cannot be convincingly explained who the real users of the dictionary are and how they are likely to benefit from consulting it, as it barely shows significant improvements on the dictionaries already existing before its compilation. What is significant though, is that this may not be regarded as the failure of the TNLU per se. The dictionary may rightfully be seen as a reflection of a lexicographic achievement of the comprehensive lexicographic process initiated through the establishment of the NLUs following the demise of apartheid. It is remarkable that without the words *Tshivenḁa/Venda* on its cover and front matter pages, the TETD is more or less similar to the *IsiNdebele/English Isihlathululi-magama Dictionary*, *Sesotho sa Leboa/English Pukuntsu Dictionary* and *Dikixinari ya Xitsonga/English Dictionary* if the names of the languages treated in these dictionaries are omitted. Not only their cover entries but also their front matter texts are identical, providing very little information about the language, the procedures followed at various stages of dictionary compilation, and, despite being called explanatory dictionaries, limited macrostructural representation and lexicographic treatment of lemmata. This is disappointing because, before the establishment of the NLUs, different needs were identified by the different language repre-

sentatives at the 1996 lexicography meeting, which logically means that the NLUs had to produce different dictionaries to satisfy the different needs experienced in the respective languages.

The above situation has possibly changed the general perception about government-planned lexicography, which was initially celebrated (Madiba 2002). This has seen the emergence of a parallel lexicographic process initiated by publishers and undertaken by free-lance lexicographers. The obvious case in point is represented by the Oxford bilingual dictionaries, namely the *Oxford Bilingual School Dictionary: Northern Sotho and English* and the *Oxford Bilingual School Dictionary: Zulu and English*, both edited by Gilles-Maurice de Schryver. These Oxford dictionaries promise to be a great success not only because of their more prestigious publisher, but also because of an unambiguous identification of their target users and efforts of meeting the needs of these target users, who are primarily school learners of either of the treated languages. Not only have the dictionaries been recognised by the Department of Education as useful language learning instruments, but the Northern Sotho and English dictionary has also won the SATI (South African Translators' Institute) award for outstanding translation dictionaries in 2009. This shows that the dictionaries have not focused on specific users and functions at the neglect of others. These dictionaries may also serve as useful translation instruments. However, it should be reiterated that while these dictionaries are generally different from and of better quality than those produced by the NLUs, the recognition of the potential of lexicography in South Africa has created a favourable climate for lexicographic practice, either by the NLUs or by commercial lexicographers. In short, the South African lexicographic processes have seen an upsurge of lexicographic practice, with dictionaries of varying quality being produced, through governmental and commercial structures.

4. Conclusion

This article has attempted to add an important and over-due dimension to the metalexigraphic account of the South African lexicographic processes. This dimension is the evaluation of the outputs of lexicographic practice since the establishment of the NLUs about a decade ago. The establishment of the NLUs remains a commendable idea which has undoubtedly improved lexicographic practice in the country. Lexicographers who previously faced different challenges as they worked on languages with established lexicographic history albeit within constrained frameworks found themselves working with the sanction of a democratic South Africa. On the other hand, formerly marginalised languages, now have lexicographers working with other language practitioners, albeit with more challenges owing to a lack of a favourable lexicographic history in their languages. The net effect is that all of South Africa's official languages have at least one dictionary. While some of the dictionaries are likely to fare unfavourably when subjected to dictionary criticism, they are certainly a

step forward in the attainment of perfect dictionaries (Gouws and Prinsloo 2005: 42). The South African lexicographic community will gradually become lexicographically informed owing to the dictionaries that have been produced over the last decade and the lexicographers will certainly benefit from their acquired experience. Ultimately, more and better dictionaries will hopefully be produced. However, for this to happen, more insights are needed into the efficiency of and challenges to the country's lexicographic establishment, especially the NLUs, so that its operations may be improved. More theoretical insights are also required to support the lexicographic practice. For example, Atkins (2002: 9) stresses that for devising tomorrow's dictionary it will be necessary to pay attention to, among other considerations, "a clear idea of its users and what they are going to do with it". This article has focused on one dictionary which seems quite problematic concerning this issue. It is evident that there is room for improvement of the available products of the South African lexicographic processes, especially given that the bilingual dictionaries such as the TETD have to be considered as a step towards the production of comprehensive explanatory monolingual dictionaries for the official languages.

References

- Alberts, M. 1999. The Importance of a Business Plan when Planning a Lexicographic Project. *Lexikos* 9: 188-197.
- Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Atkins, S. 1998. *Some Discussion Points Arising from Afrilex–Salex'98*. Unpublished Course Evaluation Document. Pretoria: University of Pretoria.
- Atkins, S. 2002. Bilingual Dictionaries: Past, Present and Future. Correard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S Atkins*. Stuttgart: EURALEX.
- Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. and S. Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Beukes, A.-M. 1996. Preface. *Lexicography as a Financial Asset in Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.
- De Schryver, G.-M. 2009. Lexikos at Eighteen: An Analysis. *Lexikos* 19: 372-403.
- De Schryver, G.-M. et al. (Eds.). 2008. *Oxford Bilingual School Dictionary: Northern Sotho and English / Pukuntšú ya Polopedi ya Sekolo: Sesotho sa Leboa le Seisimane*. Cape Town: Oxford.
- De Schryver, M.-G. et al. (Eds.). 2010. *Oxford Bilingual School Dictionary: Zulu and English / Isichazamazwi Sesikole Esinezilimi Ezimbili: Isizulu NesiNgisi*. Cape Town: Oxford.
- Golele, N.C.P. 2005. *Dikixinari ya Xitsonga / English Dictionary*. Cape Town: Pumelela.
- Gouws, R.H. 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset in Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.

- Gouws, R.H.** 2001. Lexicographic Training. Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.
- Gouws, R.H.** 2003. Towards the Formulation of a Theoretically Motivated Model for the National Lexicography Units in South Africa. Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts*: 218-246. London/New York: Routledge.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Kumalo, M.B.** 1999. The National Lexicography Units — Existing and Prospective. *Lexikos* 9: 211-216.
- Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography*. Second Edition. New York/Cambridge: Cambridge University Press.
- Madiba, M.** 2002. *Lexicography Planning in a Post-Apartheid South Africa: What Government Has to Do with it*. Paper presented at 'A Gathering of Dictionaries' seminar, organised by the Dictionary Research Centre at the University of Birmingham, 10 May 2002.
- Mafela, M.J.** 2005. Meaning Discrimination in Bilingual Venda Dictionaries. *Lexikos* 15: 276-285.
- Mafela, M.J.** 2008. L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography. *Lexikos* 18: 366-373.
- Mahlangu, K.S. (Ed.).** 2006. *IsiNdebele/English Isihlathululi-magama Dictionary*. Cape Town: Pumelela.
- Marole, L.T.** 1932. *Phrase Book for English and Venda*. Sibasa: Marole Book Depot.
- Marole, L.T.** 1955. *Afrikaans–Venda: Vocabulary and Phrase Book*. Sibasa: Marole Book Depot.
- Marole, L.T.** 1955a. *Phindulano: English–Venda Phrase Book*. Sibasa: Marole Book Depot.
- Marole, L.T. and F.J. de Gama.** 1936. *English–Tshivenda Vocabulary*. Sibasa: Marole Book Depot.
- Marole, L.T. and F.J. de Gama.** 1954. *English–Venda Vocabulary*. Third edition. Sibasa: Marole Book Depot.
- Mawela, A.** 1999. The State of Tshivenda Lexicography. *Lexikos* 9: 251-254.
- McCallum, K.** 1996. Oxford University Press. *Lexicography as a Financial Asset in Multilingual South Africa*: 121-126. Pretoria: Department of Arts, Culture, Science and Technology.
- Mojela, V. (Ed).** 2006. *Sesotho sa Leboa / English Pukuntšu Dictionary*. Cape Town: Pumelela.
- Mongwe, M.J.** 2006. *The Role of the South African National Lexicography Units in the Planning and Compilation of Multifunctional Bilingual Dictionaries*. Unpublished M.Phil. Thesis. Stellenbosch: Stellenbosch University.
- Murphy, M.L.** 1997. *Venda: CBOLD* [online]. <http://www.cbold.ddl.ishlyon.cnrs.fr/CBOLD/Lexicons/Venda.Murphy1997>. [Accessed 11 January 2010.]
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.
- Tshikota, S. (Ed.-in-chief).** 2006. *Tshivenda–English thalusamaipfi ya u thlutshedza ya nyambombili/English–Tshivenda Bilingual and Explanatory Dictionary*. Cape Town: Pumelela.
- Van Warmelo, N.J.** 1937. *Tshivenda–English Dictionary*. Department of Native Affairs. Ethnological Publications Vol. VI. Pretoria: Government Printer.

- Van Warmelo, N.J.** 1958. *Teo dza Tshivenda: Venda Terminologie — Venda Terms*. Coined for use in schools by the Venda Language Committee. Department of Native Affairs. Ethnological Series No. 39. Pretoria: Government Printer.
- Van Warmelo, N.J.** 1989. *Venda Dictionary: Tshivenda–English*. Pretoria: J.L. van Schaik.
- Wentzel, P.J. and T.W. Muloiwa.** 1976. *Drietalige Elementêre Woordeboek/Trilingual Elementary Dictionary: Venda–Afrikaans–English*. Documenta No. 16. Pretoria: University of South Africa.
- Wentzel, P.J. and T.W. Muloiwa.** 1982. *Thalusamaipfi ya nyambotharu yo khwiniswaho: Luwenda–Luwuru–Luisimane/Verbeterde drietalige woordeboek: Venda–Afrikaans–Engels/Improved Trilingual Dictionary: Venda–Afrikaans–English*. Pretoria: University of South Africa.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Terminology and Lexicography Research and Practice 9. Amsterdam/Philadelphia: John Benjamins.

The Monolingual Lusoga Dictionary Faced with Demands from a New User Category

Minah Nabirye, *Department of African Languages and Cultures, Ghent University, Belgium; Institute of Languages, Makerere University, Kampala, Uganda; and Menha Publishers, Kampala, Uganda (mnabirye@gmail.com),*
and

Gilles-Maurice de Schryver, *Department of African Languages and Cultures, Ghent University, Belgium; Xhosa Department, University of the Western Cape, Bellville, South Africa; and TshwaneDJe HLT, Stellenbosch, South Africa (gillesmaurice.deschryver@UGent.be)*

Abstract: In this article, a case is presented of an existing dictionary that is aimed at users with a minimum of primary 7 education, now faced with demands from users in primary 1–3. The reason for this demand is the result of the fact that Lusoga is currently being implemented as a medium of instruction in Uganda, in an environment where there is hardly any literature to serve the intended purpose. A review of the existing literature in and on Lusoga shows that the monolingual Lusoga dictionary — *Eiwanika ly’Olusoga* (WSG) — is the only reference work with essential information, in Lusoga, that can initiate the teaching of Lusoga at the elementary level. Although the information in the WSG may fit the purpose at hand, that information is mainly presented as a summary, with statements of conclusions only. Explanations to ease its access to the new user are thus missing. Findings from a pilot study conducted by the National Curriculum Development Centre on the implementation of the teaching of Lusoga reveal that the new user is not only the primary 1–3 pupil, but also the teacher who will need to instruct that pupil. Since children’s literature requires additional consideration beyond what can be presently availed, and since the WSG was actually compiled for an advanced user, the focus is shifted from the primary 1–3 pupil to the primary teacher. For that teacher, it is suggested to compile an additional Guide, expanding on the various extra-matter texts and especially the Language Portrait found in the WSG. This is done on the assumption that once the information is expanded and re-represented, a teacher will be able to combine the information in the Guide with that in the WSG, in order to make a Lusoga syllabus from which to draft Lusoga lessons. Although the ideal would of course be to be able to produce fully-fledged customised primers from scratch, this article’s main argument is that in the absence of both human and financial resources to do so, one can reuse and expand on the data found in an existing higher-level dictionary.

Keywords: LANGUAGE POLICY, MOTHER-TONGUE EDUCATION, LUSOGA, UGANDA, PRIMER, SYLLABUS, GUIDE, ELEMENTARY USER, PRIMARY TEACHER, TEACHER TRAINING, ORTHOGRAPHY, GRAMMAR, LINGUISTIC INFORMATION, MONOLINGUAL DICTIONARY, EXTRA-MATTER TEXTS, LANGUAGE PORTRAIT, PICTORIAL ILLUSTRATIONS

Obufunze: Eiwaniika ly'Olusoga lizuuseeku omukozesa ataali muluubilile.

Mu lupapula muno mulagibwamu engeli *Eiwaniika ly'Olusoga* elyawandiikilwa omukozesa atuuseeku mu kyomusanvu bwe lizuuseeku omukozesa okuva mu kibiina kyolubelyebelye okutuuka mu kyokusatu. Kino kiidhie lwa kuba ensomesa y'olulimi Olusoga eli kuteebwa mu nkola aye ng'ebiyetaagisa okutuukiliza omulimu guno bikaali kutuukilizibwa. Okwekeenenia kw'ebiwandiiko ebiliwo ku oba mu Lusoga ebisobola okwemelezaawo ensomesa mu lulimi Olusoga kuzwile nti *Eiwaniika* ni lyonka elilina ebyetaagibwa okusimba omusingi gw'ebiyendhegelesa mu lulimi Olusoga ebisobola okutuukana n'ebiyetaago by'omukozesa aluubililwa mu kiseela kino. Waile ng'amakobo g'*Eiwaniika* gaaliba gatuukana n'ebiyetaago by'omukozesa ono, engeli ye gaategekebwa mu *Iwaniika* teyanguyila mukozesa ono muyaaka kugeeyunila bukalamu. Okwinhonzola kw'amakobo g'*Eiwaniika* kwasinga kukolebwa mu musomo amakobo gano mwegaasibuka. Kale kyaba nti amakobo agandi gazila bulungi businziilo mu *Iwaniika* busobola kwetengelela kuyamba omukozesa aluubililwa. Ebyava mu kugezesa ensomesa y'ebynnimi enzaalilanwa mu Uganda okwakolebwa ekitongole kya National Curriculum Development Centre byazuula nti omusomi ti ni yenka ayenda okutendekebwa aye n'omusomesa agya akusomesa yeena ayenda obuyambi. Engeli ebwandiiiko by'abaana abato bwe by'ekengelwa einho ate ng'omukozesa omuluubilile ow'*Eiwaniika* yali muntu mukulu, eisila lili ku musomesa w'abasomi abo. Okutendeka omusomesa ono mu bimwetaagibwamu ng'asomesa Olusoga, akatabo akandi akalaga engeli y'okusomesaamu amakobo g'*Eiwaniika* kaidha kuwandiikibwa okumulaga engeli y'ayinza okukozesa ebili mu nnhandhula ni mu nfaanana y'olulimi mu *Iwaniika* okutegeka amasomo g'Olusoga ag'endhawulo. Waile nga kyali gwayne nti ebitabo ebite eisila ku nsomesa y'Olusoga mu pulaimale n'ebyalisnze kweetaagibwa, olupapula luno lusinziile ku mbeela eliwo buti ey'eibula ly'ensako n'abakugu abanoonheleza ku Lusoga, n'okusalawo okukozesa ekiwandiiko ekiliwo mu kiseela kino okugaziya emigaso gya kyo eli okutembeeta ensomesa y'olulimi Olusoga ng'esinziilwa ku *Iwaniika*.

Ebigambo ebikulu: EITEEKA LY'EBYENNIMI, ENDHEGELESA Y'ENNIMI ENZAALILANWA, OLUSOGA, UGANDA, ENTEGEKA Y'AMASOMO, OMUKOZESA WA P1-P3, OMUSOMESA WA PULAIMALE, ENTENDEKA Y'ABASOMESA BA PULAIMALE, WALIFU, GULAAMA, AMAKOBO GA KANALULIMI, EIWANIKI LY'OLULIMI OLULALA, EBIWANDIIKO BY'OMU BIKUGILO, ENFAANANA Y'OLULIMI, EBIFAANANI

1. Language policy vs. reality in Uganda

Lusoga is an eastern interlacustrine Bantu language spoken in Busoga, Uganda, by slightly over 2 million Basoga. The Basoga are the third-largest ethnic group in Uganda, following the Baganda and Banyankore (UBS 2006: 23). Lusoga is poorly documented and the Basoga have never studied their language. To this day the Basoga have thus never learned to read or write Lusoga. Lusoga is still categorized as an oral language.

The role of Lusoga as a medium of instruction in the region was compromised by the language policies passed in Uganda. For instance, the 1944 Makerere Conference on Language in Uganda chose Luganda as the medium of instruction in Busoga for the primary level, and in 1965 English was introduced as a subject in primary for the whole of Uganda (Ladefoged et al. 1972: 87-99). Lusoga was not considered as a medium of instruction until, in September

2005, the Ugandan Parliament passed the teaching of Lusoga as one of the nine regional indigenous languages (NCDC 2006: 5). These nine regional indigenous languages include five Bantu languages: Runyoro-Rutooro (JE11–12), Runyankore-Rukiga (JE13–14), Luganda (JE15), Lusoga (JE16), and Rukonjo (JD42); as well Sudanic and Nilotic languages: Lugbara (Central Sudanic), Acholi (Western Nilotic), Ateso (Eastern Nilotic), and Karamojong (Eastern Nilotic) — cf. NCDC (2006a: 5).

Mother-tongue education in Uganda is currently being implemented in lower primary (P1–P3) — that is, for the first three years of schooling — in spite of not having the required environment to achieve the envisaged goals effectively. For instance: Lusoga, which has featured in the language policy as a medium of instruction in the Busoga region for five years now, has a shortage of qualified mother-tongue teachers as well as a shortage of reference materials needed to sustain its new status. To the best of our knowledge, to date only one elementary primer for the first year (P1) has been compiled in Lusoga, the booklet *Idha Tusome 'Come and We Read'* by Wambi and Naigaga (2005). Material that can be used throughout primary education, that is the first three years of lower primary (P1–P3) and the next four years of upper primary (P4–P7), is non-existent. Within the current language-policy framework one could argue that material for P4–P7 is not (yet) required, but material for P1–P3 is.

One other monolingual reference work, the *Eiwanika ly'Olusoga* 'A Dictionary of Lusoga' (WSG, Nabirye 2009) exists, but it was conceived for a target user group with a minimum of P7 education: thus to be used from the last year of primary, through to secondary (both the ordinary "O" level (S1–S4) and advanced "A" level (S5–S6)), and onwards as need arises.¹ Looking back at the information provided in the WSG, one notices that it actually contains the information needed to facilitate the teaching of Lusoga in lower primary. Since the WSG was compiled for a more mature user, however, this article presents the predicament of the WSG which is currently facing genuine demands from a user not initially targeted.

2. The Lusoga literature currently available

Although the documentation of Lusoga has picked up since about a decade ago, there is still very little available. Most of the literature currently available on the local market is characterized as being peripheral. Such literature is thus considered inadequate to handle the formal demands for teaching Lusoga. At the other end, the few comprehensive and scholarly forms of Lusoga literature that exist are not readily available locally, too general and theoretical, and too advanced to serve the needs of the user who currently needs attention. The available literature may be grouped into the four categories discussed below.

The first category consists of so-called Lusoga 'grammars' (Babyale 1999, Korse 1999), booklets on Lusoga orthography (Kajolya 1990, LULANDA and CRC 2004), and Lusoga wordlists (Korse 1999a, Gonza 2007). Babyale (1999)

was undergraduate research with very basic descriptions of Lusoga. Korse (1999), on the other hand, was compiled by a non-linguist interested in enabling the pedestrian understanding of Lusoga. There does not seem to be a logical selection of the coverage and the 'grammar' is not coherent. The existing booklets on Lusoga orthography — Kajolya (1990), which is actually a revision of the much older Byandala (1963), and LULANDA and CRC (2004) — were found to be inconsistent in their description of the Lusoga orthography. The 2004 orthography, however, has the advantage that it is mostly written in Lusoga. Up until the publication of the WSG, only two wordlists had been available, both with glosses in English: Korse (1999a) and Gonza (2007), with the latter basically a reprint of the former.

The material from this first category cannot be used as reliable references for a monolingual user because this literature is largely bilingual, aimed at a bilingual user. This type of literature serves users who know English and Lusoga and bars the lower primary users from accessing it because (a) this audience does not speak the second language in addition to Lusoga, and (b) a review of the existing Lusoga literature conducted in Nabirye (2008) revealed that most of this locally-produced Lusoga literature had a very shallow coverage. For these reasons, the literature in this first category is not appropriate to teach Lusoga as a mother tongue in lower primary.

In the second category are studies in which Lusoga is featured almost coincidentally. These studies include surveys of the interlacustrine Bantu languages, where Lusoga is typically mentioned only in comparison with other Bantu languages (e.g. Tucker and Bryan 1957, Matovu 1992, Schoenbrun 1997, Matovu and Walusimbi 2000).² In most cases, the descriptions of Lusoga are generalized with a few isolated examples only in Lusoga. These studies provide findings that benefit advanced learners of Bantu languages as well as comparative and historical linguists.

The third category contains studies devoted to the description of Lusoga proper. These include three MA dissertations that are entirely dedicated to Lusoga — two of which were produced at Leyden University, in the Netherlands: Steeman (2001), in which a Lusoga play is interlinearized, and Van der Wal (2004), on Lusoga phonology. The third was produced at Makerere University, in Uganda: Nabirye (2008), in which Lusoga lexicography is described. Four scientific articles on Lusoga appeared in addition, namely Yukawa (2000), Nabirye (2009a), Nabirye (2009b), and De Schryver and Nabirye (2010). Although these studies exist, their presence is not yet fully appreciated because Lusoga is not yet used at the level on which such literature is required.

In the fourth category is literature mostly intended for the monolingual Lusoga audience at large. Works in this category include the New Testament in Lusoga (BSU 1998),³ around a dozen Lusoga story books (Lyavala-Lwanga 1967, 1969, CRC 1998, 1999, 1999a, 1999b, 1999c, 1999d, 1999e, 1999f, 1999g, 2000, 2000a, 2002, 2003), a unified orthography (Namyalo et al. 2008), and the WSG. The New Testament and the story books do not really guide the study of

Lusoga because they do not consider any Lusoga language descriptions. These works can however be utilized in addition to more specialized language-teaching material and guide towards the practice of correct grammatical structures and authentic language use. Namyalo et al. (2008) provides information about the standard writing system for Lusoga.⁴ Lastly, the WSG being the first explanatory dictionary for Lusoga, this meant that it specified foundational language information in Lusoga for the very first time in order to complete the relevant lexicographical tasks. As it turns out, this is the type of information presently required for the formalization of Lusoga as a medium of instruction.

References in categories 1–3 have handled selective descriptions of Lusoga for bilingual audiences or addressed advanced learners, and have neither specified the most basic but essential language content nor its relevant formalized terminology in Lusoga. Lusoga teachers for primary and secondary schools who have to use Lusoga as the medium of instruction find it virtually impossible to successfully carry out their job with the literature mentioned in categories 1–3. Moreover, primary teachers will obviously find it very difficult to read through and formulate a Lusoga syllabus from all the resources that do not specify Lusoga language content and the respective terminology needed to facilitate their lessons.

We can conclude from this overview that the specification of the structure of Lusoga has not yet been given independent scrutiny. Namyalo et al. (2008) and the WSG are the two references that have information that can be directly utilized for purposes of teaching the Lusoga language, in Lusoga, in lower primary. This is so because both of these references provide essential information regarding the structure of Lusoga, information which is needed for the instruction of Lusoga in formal settings. However, Namyalo et al. (2008) has the shortcoming that the metalanguage is English, making it difficult for monolingual users to access. The WSG thus stands out as an ideal Lusoga reference because it specifies all its information in the language of the user. The WSG, then, seems to be the only reference currently available which provides some of the most basic types of information about the structure of Lusoga, information that can serve the purpose of teaching Lusoga to mother-tongue audiences. We will now move on to see the kind of information types described in the WSG that can sustain the study of Lusoga at an elementary level.

3. Language information specified in the WSG

In the few studies prior to Namyalo et al. (2008) that handled the rules of spelling in Lusoga, gaps existed. For instance, the specification of full word forms and the different types of words in Lusoga had not yet been established. This made it difficult to demarcate parts of words and consequently hampered the specification of the right spellings. The writing of complex words (i.e. the writing of compounds) and words of foreign origin was problematic too, mainly because the existing orthographies at the time lacked the ability to cater

for some of the irregularities that arose in their specification. The alphabet was found to be unjustifiably restricted.

The WSG set out to describe the letters of the Lusoga alphabet, with justifications for the new additions, to provide a basis for the writing system of Lusoga. Findings from this research on the specification of the Lusoga alphabet and rules of writing were subsequently tabled in the process of harmonizing the orthography for eastern interlacustrine languages in Uganda and published as Namyalo et al. (2008).

Spelling was also addressed when deciding on the citation forms to be used in the WSG. The research on the morphology of Lusoga led to the formulation of citation styles different from the citation styles found in earlier monolingual Bantu dictionaries. In order to apply the new citation style which uses full word forms (cf. Nabirye 2008, Nabirye 2009a, Nabirye 2009b),⁵ words falling in the verb category especially, were further segmented down to their smallest meaningful parts (morphemes) within the entry. All types of entries and entry styles were described and explained in the front matter of the WSG. These two methods (full words and word segmentation at the lemma level) paved the way for the specification of the Lusoga orthography currently in use.

The second major undertaking for the WSG was the allocation of parts of speech to over 12 000 entries. Although both Korse (1999a) and Gonza (2007) are bilingual bidirectional English–Lusoga wordlists, with the word 'dictionary' misleadingly in their titles, they only allocated parts of speech to some of the English lemmas, but not to any of the Lusoga lemmas. See Addenda 1–3.

Neither the Lusoga terms for the different word classes nor a list of Lusoga words with word-class indications existed in the previous references for Lusoga. Specification of this information was therefore pioneered in the WSG. A list of all the word classes that were coined for Lusoga is given and illustrated in the front matter of the WSG, and all the lemmas in the WSG carry a part-of-speech label. See Addenda 4–5.

The WSG also allocated gender to Lusoga nouns and perfective forms to verbs, with a special treatment for irregular forms. All irregular forms were entered and cross-referenced to the entries from which they are derived. In some cases, irregular class genders were written out in full and included with the respective entries to help highlight such forms for the user. Treatment of this nature was also extended to compounded noun forms which are a result of gemination or sound-change rules. Users were either given the derived forms or the canonical forms with a cross-reference to the relevant base entries. Examples of irregular entries are given below (with cl. = class).

- | | | |
|-----------------------|-----------|---|
| (a) (a)kadaada | <i>l.</i> | [12/14, 9 (e)ndaada] |
| type of bird | | cl. 12 & 9: type of bird / cl. 14: type of birds |
| (a)kagina | <i>l.</i> | [12/14, 3 (o)mugina] |
| lice egg | | cl. 12: lice egg / cl. 14: lice eggs, cl. 3: collection of lice eggs |

| | | |
|-----------------------|------------|--|
| (b) Kyabazinga | <i>l.</i> | [1/2 Bakyabazinga] |
| King | | cl. 1: King / cl. 2: Kings |
| kyapati | <i>l.</i> | [9/10 dhikyapati] |
| type of food | | cl. 9: type of food / cl. 10: type of foods |
| (c) (o)ku.b.a | <i>t.</i> | [- <i>baile</i>] |
| to be | | was |
| (o)ku.biiw.a | <i>kt.</i> | [- <i>biiye</i>] |
| to go/become bad | | went/became bad |
| (d) nnonze | <i>kt.</i> | Bona: (o)ku.lond.a |
| I have picked | | See: to pick |
| (e)mpya | <i>l.</i> | Bona: (o)luya |
| courtyards | | See: courtyard |
| (e) (e)mmamba | <i>l.</i> | [9/-, 12 (a)kabamba] |
| meat | | cl. 9: meat, cl. 12: small piece of meat |
| (e)nnunga | <i>l.</i> | [9/10, 12 (a)kalunga] |
| black ant | | cl. 9 black ant / cl. 10: black ants, cl. 12: small black ant |

Specification of the concepts for the most common and useful grammatical and linguistic terms in Lusoga was also one of the tasks undertaken in the WSG for the very first time. The terms specified include: Lusoga words for punctuation marks (i.e. full stop, question mark, quotation mark, etc.), parts of speech (i.e. noun, verb, etc.), linguistics terms (i.e. morphology, syntax, etc.), grammatical terms (i.e. gender, aspect, mood, affixes, enclitics, etc.), usage labels (i.e. argot, ameliorative, literal, private, etc.), and symbols (i.e. backslash, asterisk, etc.). The contribution rendered to a language by specifying information of this nature may seem trivial, especially in well-documented languages, but this information did not exist for Lusoga and had to be provided for the very first time in the WSG. This was an important contribution to the essentials needed in order to set a standard for the formalization of Lusoga as a language of instruction.

The main shortcoming to the specification of this information in Lusoga is that since Nabirye (2008) was an academic study (in the form of an MA dissertation), the detailed discussions on how the specification process was carried out as well as the results are mainly treated in the study itself. Some of the results are therefore merely 'used' in the WSG, with no explanation on how the conclusions were arrived at. Other results, although obtained through the same research process, were considered too new and needed further consideration before being adopted, so those were only 'mentioned' in the WSG. As such, the WSG did not include entries and definitions for all the newly-specified Lusoga language terminology. All grammatical and linguistic data accrued from the research process is only summarily interspersed throughout the WSG, with the main objective of attaining lexicographical goals.

The intended user of the WSG was assumed to have attained some formal

grounding and to have the ability to read and understand the information in the WSG. Though the elementary school category was not directly addressed, it now turns out that this group most urgently needs to use the WSG. Yet this user category needs more assistance on how to decode data in the WSG. A new and more focused presentation style of the WSG information is therefore in order to address the new target audience.

4. Mother-tongue education in Uganda

After the Ugandan Parliament passed the teaching of nine regional indigenous languages of Uganda as a medium of instruction in lower primary, the implementation of this language policy was passed on to the National Curriculum Development Centre (NCDC). In the period 2005–2006, the NCDC embarked on the development of the so-called 'thematic curriculum' and syllabus for the teaching in the mother tongue in primary 1–3, which was passed in July 2006 by the Minister of Education and Sports, to be implemented starting in February 2007 (NCDC 2006: 1). The thematic curriculum and syllabus was piloted in the districts of Kampala, Rakai, Rukungiri, Kasese, Kabarole, Arua, Gulu, Moroto, Kumi, Iganga, and Nakasongola (NCDC 2006: 1). The pilot districts were chosen on the basis of their coverage of the nine regional indigenous languages that met the conditions for approval.

During the testing stage, however, it was observed that teachers in most of the regional indigenous languages were not using those languages in instruction, even though they were time-tabled. Another observation was that the teachers supposed to teach these languages were instructed and trained in methodologies for teaching *foreign* languages, and lack the right methodology and know-how for the formal instruction of *indigenous* languages. Moreover, the teachers themselves do not know the indigenous languages fully. Even though they may be good speakers, they cannot read or write or ably teach the indigenous languages. It was noted that this is further hampered by the fact that some of the languages, amongst them Lusoga, largely exist orally and are not documented. Instructional materials like dictionaries, textbooks and readers are lacking.

4.1 Lusoga as a medium of instruction in primary 1–3

In Busoga, the pilot initiative of teaching in Lusoga was carried out in the districts of Iganga and Namutumba, in fifteen primary 1–3 schools (CRC and CCFU 2008: 1). In late 2006 and early 2007, the Cultural Research Centre (CRC), with the financial and technical support from the Cross-Cultural Foundation of Uganda (CCFU), carried out research in these districts to gauge the impact of the introduction of Lusoga as a medium of instruction in lower primary classes. The main advantage noted in this pilot was that there was improved perform-

ance and student participation in class activities. There also was a noted renewed sense of self-value.

However, the CRC observed that sustainability of the indigenous-language teaching programme had many challenges, mainly because teachers had a relaxed attitude towards the teaching in and of Lusoga which was regarded as meant for the poor and boring. Also, Luganda and English were preferred as language subjects because they are formalized and are easier to teach when compared to Lusoga. The latter is of course due to the lack of reference materials and a formalized foundation for the teaching of Lusoga, which raises scepticism as to the success of the language policy in the region.

4.2 Lusoga as a subject in senior 1–4

The NCDC language-teaching syllabus for the ordinary (O) level of secondary school (i.e. senior 1–4) focuses on the teaching of vocabulary, language structure, and language skills. In the case of Lusoga, a monolingual dictionary of Lusoga to facilitate the learning of Lusoga vocabulary is now available, namely the WSG. However, the references for teaching the Lusoga language structure are not yet available. Also, the readers to enable a good grounding in the learning and practice of more advanced grammar and different language skills are limited to story book primers (i.e. those mentioned in the fourth category of Lusoga references) and informal oral literature. Therefore, while the NCDC thematic syllabus has a provision for the teaching and examination of a grammar component, with respective descriptions of the teaching content at every level (NCDC 2003: 9-43), the teaching of Lusoga grammar lacks the very foundational requirements — that is standard grammar texts — to successfully carry out this aim.

4.3 Escaping the conundrum

Sections 4.1 and 4.2 sketched a grim picture: Teachers in lower primary typically obtained senior 4, yet those teachers were not given the tools to teach the regional indigenous languages (cf. §4.2), while the lower primary pupils do not have any material to study those regional indigenous languages either (cf. §4.1). Notwithstanding these problems, Lusoga *is* currently being 'used' as a medium of instruction in lower primary. Monolingual literature to serve as references is thus urgently required, and the NCDC is outsourcing for assistance to further the goal of teaching mother tongues in Uganda. It is an understatement to say that the NCDC is facing problems with the implementation of this goal, and it will not surprise anyone that they have called on studies like the WSG to assist in guiding the process of mother-tongue education in the Busoga region — and hoping for similar studies to emerge for the other regions.

5. The new target audience for the WSG

There are two kinds of users who have emerged from the findings of the NCDC pilot studies, namely the primary 1–3 pupil and his/her teacher (who typically reached senior 4).

The first user is the primary 1–3 pupil whose education system for the first three years is conducted entirely in Lusoga. This is an elementary user who is just being introduced to the formal education system. This user is fully monolingual, especially if not raised in an urban setting. He/she is just learning how to read and write and has no prior knowledge of the structure of languages. This user is also still in the process of mastering the mother tongue and needs a lot of practice in order to acquire the language in all its facets.

The second user is the primary Lusoga language teacher who is supposed to use Lusoga as the medium of instruction in primary 1–3 and as a subject in primary 4–7. The teacher in this category has only attained senior 4. He/she was trained to teach the English language and other subjects in English in primary education, but not to teach Lusoga, nor in Lusoga. This teacher has the innate structure of Lusoga which he/she displays orally but has most probably never read or written in Lusoga at all. He/she can be assumed to have good oral skills of the mother tongue but does not formally know the rules of the language he/she speaks.

From the 1950s up until the early 1980s, there was a provision for mother-tongue education in each region. Students in the Busoga region would in this instance have learned the structure of Luganda (not Lusoga!), and how to read and write it, but that provision is no longer available. For our current teacher, teaching Lusoga is boring because there is nothing extra he/she can add to the formalized mother-tongue education system other than correcting mistakes noted in speech with rules he/she has internalized but cannot explain.

Given this state of affairs the teacher and the pupil are just as 'clueless' with regard to the structure of Lusoga that they both need assistance. Presently, this teacher has resorted to recounting traditional stories in the Lusoga lessons, with no reference to the relevant structural language content. Although oral literature, which is acquired informally, can inform some teaching tasks, at the moment it lacks the formalized structure needed to facilitate the learning of Lusoga language rules. In the long run, the stories run out and the subject becomes boring to teach for the teacher, boring to learn for the pupil, and is consequently abandoned as shown in the findings of the pilot study.

The nature of the user in the elementary category is delicate. These pupils mainly learn Lusoga from informal settings where no formal structure is stipulated. Although teachers of these pupils are mature, they too lack the formal training needed to understand the Lusoga language structure and the right teaching methods to pass it on to their pupils. The teachers do not (yet) know how to decode the information in the highly-specialized academic studies listed in the second and third categories of Lusoga references. They may also

fail to comprehend or even apply the information in the WSG for their teaching purposes. In addition, one should of course also keep in mind the very real possibility that they have hardly had any exposure to dictionaries in their lives, dictionaries in any languages, and that they do not know how dictionary information could inform their teaching tasks.

In this context both the pupil and the teacher need guidance on the structure of Lusoga. The teacher needs assistance and training on what the structure of Lusoga entails before he/she can teach. He/she is most likely to benefit from the dictionary data in the WSG and by doing so improve his/her formal knowledge of Lusoga. This teacher can be guided further on the linguistic data given to enable him/her to develop a syllabus or lesson plan with the data given in the dictionary. Success of this assumption is based on the kind of information provided in the Guide to help direct the teacher's focus on the selection of topics to teach in each Lusoga lesson.

6. The WSG and the development of language training materials

In order to develop language training materials for schools in Uganda, the NCDC needs assurance that the literature in question has been prepared in accordance with the requirements set in their O-level syllabus (NCDC 2003). To understand the way in which the NCDC works, it is revealing to now briefly look at its dealings with the general public on the procuring of study materials for schools.

The Ministry of Education and Sports (MoES) in Uganda is charged with ensuring the success of the national language policy. The MoES mandates the NCDC to implement actions to be taken to ensure that the national language goals are met. The NCDC seeks the specialist help from institutions of higher learning, such as the Makerere Institute of Languages, Language Boards (these are presently instituted by the Constitution but not yet fully systematized nor functional), as well as any interested parties. Whenever necessary and on an ad-hoc basis, the NCDC also calls on religious organizations (e.g. the CRC), non-governmental organizations (e.g. the CCFU), teachers, researchers, and publishers. All of these have so far constituted the consultants on mother-tongue education for the NCDC. Though some Ugandan languages are better documented than others and consequently have more reference materials than others, skilled personnel to implement mother-tongue education in all of the nine regional indigenous languages is still lacking.

Since the government is not offering funding to interested parties to invest in research or to train mother-tongue teachers and researchers, it is not surprising that the results from the pilot study conducted by the NCDC showed that there were gaps that may necessitate the revamping of the whole programme. This is the level at which adaptation of the WSG is being suggested to help in the teaching of Lusoga. The WSG was developed at an institution of higher learning and, in addition to the advantages and contributions already

mentioned in this article, it was very well received by the general public. For a sample of the reactions to the arrival of the WSG — all unanimously positive — see for example Aryatuha (2010), Ssejjengo (2009), Tendo (2010), Tumusiime (2010), or Wandera (2010).

7. What is at stake for the WSG in this context?

The language information in the WSG, though relatively basic, is not yet truly fitting as a teachers' guide. This information needs a new presentation style befitting primary school education. In this context, the primary teacher is the user who should be addressed. This is so because, at present, this teacher's needs are easier to address compared to the needs of his/her pupils. The teacher mainly needs assistance on how to develop a syllabus and lesson plans for teaching Lusoga, using the data that already exists in the WSG. The main goal in the remainder of this article is therefore to specify the relevant WSG information for this user, and to show how lessons can be drawn from it. For this teacher, two monolingual Lusoga references will eventually be availed, namely the Teachers' Guide and the WSG.

Though the results of the Nabirye (2008) study are undoubtedly a contribution to the study of Lusoga, they have their own shortcomings. Chiefly, not all findings from this study have so far been tested in full on the target audience. Although the way in which the WSG data is presented may seem prescriptive, directing users to 'the right way' Lusoga ought to be used, it is a compromise that lexicographers in this position have to contend with. Recall Zgusta's caution to lexicographers in this regard: "anticipate that [your decisions] will not be to everybody's liking and that the real development of the language itself may decide against [your] choice" (1971: 293). The specification and description of language information in the WSG was however based on a logical analysis, therefore any issues arising for or against the findings of the study are open to discussion.

The advantage, of course, is that procuring a proper testing measure for the WSG at this point in time, something initially thought to be a dream, could be born out of this process. Users will place more demands on the dictionary and strive to understand the information, so that hopefully more questions will arise about the WSG data. The new proposed Lusoga language terms now have a chance to be tested to see whether they are appropriate, adequate and understood by the target audience.

8. The primary Teachers' Guide for the teaching of Lusoga

At present, the authority of a Teachers' Guide is what the NCDC would like to have in order to consider the literature worthy of a reference for teaching Lusoga. The information specified for the Teacher's Guide and how it is to be presented is given in the sections that follow.⁶

It is important to note that the primary teacher training course does not restrict trainees to the teaching of certain sections only of the primary school system, thus either lower primary or upper primary. Consequently, although this article aims to address the teaching of Lusoga in lower primary, it will also specify Lusoga data for primary 4–7 wherever applicable. In the upper primary level, Lusoga is taught as a subject in Lusoga (as is the case for lower primary), although the medium of instruction for everything else is English. Coverage of information necessary for the teaching of Lusoga in the entire primary section helps the primary teacher to get an overall orientation on how to conduct Lusoga lessons in whichever level he/she ends up teaching.

The subsections that follow summarize topics to be covered based on the information given in the central section of the dictionary, the extra-matter texts, as well as the Language Portrait found in the WSG. These topics include the orthography, pictorial illustrations, word categories, the counting system, quantification and measuring, usage labelling, a language portrait, and a glossary of terms. A summary of the contents in each section is given with sample references on how the topics can be actualized in a lesson.

8.1 The orthography

In this section, reference will be made to Namyalo et al. (2008) because it is now the orthography approved by the NCDC as the standard for writing Lusoga. A detailed specification of the new alphabet and the reasons why earlier studies have deviated from this should be covered. This discussion will be based on the justifications found in the WSG. For example, letter 'C' featured in the WSG alphabet to host abbreviations, but is not used otherwise in Lusoga. Letter 'J' was also introduced to cite foreign words that have entered into Lusoga and are now in standard usage, such as *Jinja* (a major town in the Busoga Kingdom, as well as the name of the district around it). Letter 'R' was dropped from the Lusoga orthography based on the argument that it is not the sound rendered in speech by the Basoga; it was only maintained in the WSG to cater for abbreviations. Clarification of this nature is of importance to the teacher and aims to show that the alphabet can be dynamic and altered with justifications to serve specific purposes as was the case in the WSG.

Teaching the rules of writing should be introduced in lower primary. This can be achieved by teaching the writing of single, double, and triple letters first. Exercises testing pupils on words with contrastive spellings, or thus for example those having short vs. long sounds, can also help the teaching of the letters used in the Lusoga alphabet. After that the focus should be on the teaching of Lusoga spelling. Pupils can for example be asked to fill in missing parts of words or sentences not only to test their competence in writing, but also to (implicitly) test their knowledge of morphology and vocabulary.

Teaching the composition of sentences and the right word order should also be introduced in lower primary. This will be coupled with teaching the use

of the right punctuation marks within and at the end of sentences. A list of all the Lusoga terms for the punctuation marks is given in the Language Portrait of the WSG. Teaching of punctuation can actually be an independent lesson. The teaching of capitalization for proper nouns (personal names and place names) and for words at the start of sentences is also one of the ways to teach punctuation in lower primary. Exceptions to the rules of writing and sound changes can be introduced later, in upper primary, and given further attention in secondary school lessons of Lusoga to build on what was already learned.

Symbols used to write and represent linguistic information are given in the front matter of the WSG. All the symbols used also have their Lusoga terminology specified for the very first time in the WSG. Examples of symbols are punctuation marks, symbols for the formalized representation of linguistic information, symbols found on a keyboard, etc. The different types of symbols serve various tasks and may not all be important for the primary section. However, some are really vital like the punctuation marks which should be featured in relation to the proper writing of sentences. As pointed out, this is a subject that should be introduced in lower primary. A selection of the more advanced symbols for the formalized representation of linguistic information, such as phonetic symbols, may be introduced in lower primary too to teach the pronunciation of some letters of the alphabet and words with contrastive sounds. This can later be expanded in upper primary to introduce dictionary awareness as an integral part of language learning. Exercises at the upper primary level could involve studying what constitutes dictionary information and how to find words plus meanings in dictionaries.

The alphabet, rules of writing, pronunciation, punctuation, and dictionary use are subjects that can feature in primary lessons of Lusoga. All these subjects help in the teaching of the Lusoga orthography. The same subjects can then be expanded as the teacher wishes to teach specific language skills.

8.2 Pictorial illustrations

Throughout the WSG, there are hundreds of drawings — illustrating various single lemmas and bringing exponents of certain semantic fields together — and these may be singled out. On the one hand such drawings can be used to mimic a picture dictionary, a type of dictionary that would otherwise be typical for lower primary users, and on the other hand these drawings can be used to develop lessons that revolve around vocabulary building. In both these scenarios, the links back to the lemmas in the central section of the dictionary itself should be kept in mind for upper primary, as doing so constitutes a gentle approach to creating a genuine dictionary culture.

For example, Addendum 6 shows the pictorial illustration found at the lemma *egaali* 'bicycle'. Not only is the object of the lemma itself depicted, but all the important parts of that object are named as well. Needless to say, all those parts have been included in the WSG as lemmas in their own right.

As another example, Addendum 7 shows one of the eight full-page plates with illustrations included in the middle matter of the WSG, in this case for various modes of transport (not repeating those already covered at the lemmas themselves). Other onomasiological fields found in the WSG include fruit and vegetables, animals, insects, musical instruments, clothes, etc.

8.3 Word categories

Terms specified for the Lusoga word categories are given and exemplified in the front matter of the WSG. A description of the nature of information in each category is given but the terms are not defined in the dictionary. In the Teachers' Guide, therefore, the meaning and function of each term will be given and also treated in the glossary to be appended to the booklet. This will not only provide the first explanatory record for this type of information but also provide a testing ground for the new terms. If the proposed terms are easily understood by this audience then they can be adopted and passed as the standard Lusoga referents for word categories in future texts.

Characteristics of each word class can also be described in a series of lessons, some of them linked to the types of words shown in the pictorial illustrations (cf. §8.2), for example to illustrate the difference between nouns and verbs. The depth of teaching this subject will depend on the level of the learner. In lower primary, for example, pupils may learn the regular noun classes and regular verbs with simple tenses. In upper primary, irregular forms of words can be introduced, with possibly links to rules for sound changes.

Word formation processes can also be introduced while teaching word classes, thus focusing on full word forms, affixes, tenses and compounds, plus the proper way in which all of these should be used and written. Exercises can be varied depending on the level of the learners.

8.4 The counting system, quantification and measuring

The Lusoga counting system, as in any other Bantu language, is both a highly familiar and linguistically interesting category to spend ample time on with young learners. Even just to count from one to ten, one already uses a number of different word classes and grammatical constructions, and these language processes can be seized on for teaching purposes, linking them to noun, adjective and verb bases, as well as to the wider grammar.

Quantifying and measuring items are two further fertile topics. The teaching of age ranges, for example, is useful to vocabulary-learning lessons. In such lessons pupils can be tasked to learn terms used to refer to people in given age groups from infancy to maturity. See Addendum 8. Lessons about days, times and time durations — see Addenda 9, 10, and 11 respectively — can all be used to test the writing skills of pupils, where they detail activities of their

day. The teaching of weights (i.e. heavy vs. light; kilos, tonnes, etc.) can be used to dramatize language lessons with pupils acting out roles of selling commodities of different quantities and sizes.

Terms proposed for each of these quantification and measuring sequences will be listed in the glossary for easy reference. For those that were not already in the Lusoga system, new words were coined, and these will also have to be explained. Unlike the parts of speech, punctuation marks, and symbols, most of the terms for this category were already in the language but had not been given prominence. This is thus also one of the WSG types of data that still need to be tested. New or alternative terms not initially targeted in the WSG may arise out of this testing process and adopted in future references.

8.5 Usage labelling

A list of all the usage labels used in the WSG is given in its front matter. This data is also specified for the very first time in Lusoga but the form in which it is given in the WSG fits a more advanced user. The concept entailed in usage labelling can however be simplified and introduced in lower primary.

Exercises meant to teach usage in lower primary could focus the teaching on how sentences render emotional effects (i.e. surprise, anger, pity, etc.). In upper primary, teaching of usage could for example focus on the writing of formal vs. informal letters — examples of which are also included in the WSG middle-matter texts. Treatment of the different writing styles is a good exercise to test pupils on how to write and on how to use words properly.

Usage-labelling information can also help to test the writing and speaking skills, whereby pupils are asked to act out or write about different emotions. This exercise can help to build their vocabulary since they are tasked to choose the right words to bring about a desired emotional effect. Exercises on usage can therefore feature in lessons at different levels of education. For advanced learners this topic is essential, especially for those who would like to use Lusoga for special purposes. Lessons on usage can further be moulded depending on the level of learners and the intended objectives.

8.6 A language portrait

Addenda 8 to 11 are samples from the first two pages of the Language Portrait in the WSG. That section contains a full 15 pages (C1–C15), and is actually a goldmine of topics that can serve as starting points to build out the Teachers' Guide. Three more possible topics follow.

Gender information can for example help to explain how to differentiate between femininity and masculinity. Lessons on this subject can be featured in the teaching of parts of words where pupils are asked to mark the right prefixes that represent specific genders on names and titles. For instance, *Is-* is a

masculine prefix mostly used at the beginning of titles, while *Inh-* is the feminine counterpart. Examples of such words are *Isebantu* 'King' and *Inhebantú* 'Queen'.

Names of colours in Lusoga can be taught in the study of adjectives and story writing, where pupils could be asked to write about their colour preferences or the colouring processes they have observed both in traditional and modern life (e.g. the dying of hair vs. the dying of palm leaves for making mats or clothes). Learning about working with colours could also be further developed in other subjects such as in the study of pollination or in art lessons where pupils learn how to shade.

Lastly, coverage of the family tree and the related terms can provide pupils with interesting ways to describe their families and their relations to different people in the family structure — either from the maternal or paternal side of the family. This information also forms a bridge between school activities and home activities. At this level the parents can also be involved in the learning of Lusoga and the Busoga culture.

8.7 A glossary of terms

The glossary will consist of the new Lusoga language terms and references to other Lusoga works. The terms included will *inter alia* cover grammar, word categories, usage labels, punctuation marks, symbols, quantification, and time measurement referents. These terms will have very short definitions. This is done in order to make the Guide available as an independent reference.

9. Concluding remarks

Coverage of the Lusoga orthography and word classes in particular are considered relevant and apt for learning in primary. In later school years, a pupil can build on the foundations given during this period to obtain a broader view of the language. Coverage of efficient language usage, the counting system, and other language-related issues are also considered essential and complimentary to the learning of the two main focus areas in primary-language education. For all these areas, the WSG already contains the necessary information, albeit in synthesised form, which is of course in line with lexicographical practice. Note that this does not necessarily mean that the data given in the WSG is comprehensive in that it already covers all topics required for primary-language education. Rather, the WSG is currently the only monolingual resource on Lusoga which contains such divergent types of information, and which can thus be targeted to assist in the teaching of Lusoga at the elementary level.

The subject of this article could of course also have been addressed by means of the production of two very specific forms of literature, namely a children's grammar and a children's dictionary, both designed and customised for

the primary level. Such an approach was not suggested for the simple reason that the teaching of Lusoga is currently already being implemented, and such literature would thus be needed as of yesterday. Producing these new types of Lusoga teaching materials should of course be considered, but only in the knowledge that creating these materials from scratch will take a considerable amount of time. Moreover, funding for the national language policy is extremely poor at present, and most if not all of the documentations for mother-tongue education are currently privately sponsored. For this reason, publishers and independent researchers who wish to make some financial gain in this market are very reluctant to forge forward. That is the main reason why we have chosen to capitalize on what is already available for Lusoga, namely the *Eiwanika ly'Olusoga*, and to empower this monolingual dictionary so it acquires a more divergent application than the one originally intended.

Endnotes

1. Uganda has a 13-year education cycle: seven years for primary (P) and six years for secondary (S). Each is divided into two levels, three years of lower primary (P1-P3) and four years of upper primary (P4-P7), followed by four years of ordinary "O" level (S1-S4) and two years of advanced "A" level (S5-S6). Learners drop out at any of these levels, with very few moving on to tertiary education.
2. Although Matovu and Walusimbi (2000) is listed here, the research referred to was abandoned, and the results remain unpublished.
3. This work incorporates the earlier BSU (1996), which in turn incorporates the even earlier BSU (1994).
4. On 8 February 2010, the Centre for Advanced Studies of African Society (CASAS) signed a memorandum of understanding with the National Curriculum Development Centre (NCDC) whereby the CASAS harmonized orthographies would henceforth form the NCDC basis for standardizing the entire mother-tongue writing system. As a result, the orthographies in Namyalo et al. (2008) have now become the standard for writing Luganda, Lusoga, Luma-saaba, Lusaamya, Lulamogi, and Lunyole. M. Nabirye was responsible for Lusoga.
5. For bilingual equivalents involving other Bantu languages — in dictionaries for Cilubà, Swahili, Northern Sotho and Zulu — see De Schryver (2008, 2008a, 2008b).
6. Although English is used in this article to summarize the items covered in the guide to the teaching of Lusoga, the Teachers' Guide itself will of course be written exclusively in Lusoga.

References

- Aryatuha, M. 2010. Book Review: *Eiwanika ly'Olusoga*. *Red Pepper*, March 2010.
- Babyale, S.C. 1999. *Gulama w'Olusoga Omukalamu* [The Proper Lusoga Grammar]. Unpublished B.A. Dissertation (written in English). Kampala: Makerere University.
- BSU. 1994. *Mariko. Omutwe Ogwokuna n'Ogwokutaanu mu Lusoga* [A Selection from St. Mark's Gospel, Chapters 4 and 5, in Lusoga]. Kampala: The Bible Society of Uganda.

- BSU. 1996. *Mariko Amawulire Amalungi mu Lusoga* [The Gospel of Mark in Lusoga]. Kampala: The Bible Society of Uganda.
- BSU. 1998. *Endagaano Empyaka* [The New Testament]. Kampala: The Bible Society of Uganda.
- Byandala, G.I.** 1963. *The Lusoga Orthography*. Iganga.
- CRC. 1998. *Kintu*. Jinja: Cultural Research Centre.
- CRC. 1999. *Ababita Ababiri*. Jinja: Cultural Research Centre.
- CRC. 1999a. *Akatabo Akasooka ak'Enfumo edh'Abasoga*. Jinja: Cultural Research Centre.
- CRC. 1999b. *Amagezi Tigamalwayo*. Jinja: Cultural Research Centre.
- CRC. 1999c. *Ensambo edh'Abasoga*. Jinja: Cultural Research Centre.
- CRC. 1999d. *Mwidhe Tufume*. Jinja: Cultural Research Centre.
- CRC. 1999e. *Obufunvu Magezi*. Jinja: Cultural Research Centre.
- CRC. 1999f. *Omuwangano mu Busoga*. Jinja: Cultural Research Centre.
- CRC. 1999g. *Twire ku Butaka*. Jinja: Cultural Research Centre.
- CRC. 2000. *Enkabi n'Ekifini mu Busoga*. Jinja: Cultural Research Centre.
- CRC. 2000a. *Lwaki Abakazi Tibabeeda Mulambo*. Jinja: Cultural Research Centre.
- CRC. 2002. *Ebikoiko eby'Abasoga*. Jinja: Cultural Research Centre.
- CRC. 2003. *Ebindi kw'idembe ery'Obw'omuntu mu nsi Yoonayoona*. Jinja: Cultural Research Centre.
- CRC and CCFU. 2008. *Reclaiming our Language through Education. Teaching in Lusoga in Primary Schools: What Change In and Out of the Classroom?* Jinja: Cultural Research Centre & Kampala: Cross-Cultural Foundation of Uganda.
- De Schryver, G.-M.** 2008. Why Does Africa Need Sinclair? *International Journal of Lexicography* 21(3): 267-291.
- De Schryver, G.-M.** 2008a. A New Way to Lemmatize Adjectives in a User-friendly Zulu-English Dictionary. *Lexikos* 18: 63-91.
- De Schryver, G.-M.** 2008b. The Lexicographic Treatment of Quantitative Pronouns in Zulu. *Lexikos* 18: 92-105.
- De Schryver, G.-M. and M. Nabirye.** 2010. A Quantitative Analysis of the Morphology, Morphophonology and Semantic Import of the Lusoga Noun. *Africana Linguistica* 16: 1-45.
- Gonza, R.K.** 2007. *Lusoga-English Dictionary and English-Lusoga Dictionary*. (Revised edition of P. Korse's (1999a) dictionary.) Kampala: MK Publishers.
- Kajolya, J.B.N.** 1990. *The Lusoga Orthography*. (Revised edition of G.I. Byandala's (1963) Lusoga orthography.) Jinja: Lusoga Ecumenical Committee.
- Korse, P.** 1999. *A Lusoga Grammar*. (Two subsequent editions were published in 2000.) Jinja: Cultural Research Centre.
- Korse, P.** 1999a. *Dictionary Lusoga-English/English-Lusoga*. Jinja: Cultural Research Centre.
- Ladefoged, P., R. Glick and C. Criper** (with an introduction by C.H. Prator and additional material by L. Walusimbi). 1972. *Language in Uganda*. Ford Foundation Language Surveys 1. London: Oxford University Press.
- LULANDA and CRC. 2004. *Empandiika y'Olulimi Olusoga Enkalamu/Standard Lusoga Orthography*. Jinja: Lusoga Language Authority.
- Lyavala-Lwanga, E.J.** 1967. *Endheso dh'Abasoga*. Kampala: Milton Obote Foundation.
- Lyavala-Lwanga, E.J.** 1969. *Kiyini Kibi*. Kampala: Milton Obote Foundation.
- Matovu, C.N.** 1992. *A Synchronic Description of Lusoga in Terms of its Relatedness to Luganda*. Unpublished Ph.D. Thesis. Kampala: Makerere University.
- Matovu, K.B. and L. Walusimbi.** 2000. *A Linguistic Survey of the Current Status of the Dialects of Some Eastern Bantu Languages*. Unpublished Manuscript. Kampala: Makerere University.

- Nabirye, M.** 2008. *Compilation of the Monolingual Lusoga Dictionary*. Unpublished M.A. Dissertation. Kampala: Makerere University.
- Nabirye, M.** 2009. *Eiwanika ly'Olusoga. Eiwanika ly'aboogezi b'Olusoga n'abo abenda okwegwa Olusoga* [A Dictionary of Lusoga. For Speakers of Lusoga, and for Those Who Would Like to Learn Lusoga]. Kampala: Menha Publishers.
- Nabirye, M.** 2009a. Compiling the First Monolingual Lusoga Dictionary. *Lexikos* 19: 177-196.
- Nabirye, M.** 2009b. Dictionary Compilation for Mother-tongue Speakers of Bantu Languages. Zhu, R. (Ed.). 2009. *Proceedings of the International Seminar on Kangxi Dictionary and Lexicology*: 597-607. Beijing: Beijing Normal University.
- Namyalo, S., L. Walusimbi, G. Bukenya, M.W. Masakala, M. Nabirye and F. Kiingi.** 2008. *A Unified Standard Orthography of Eastern Interlacustrine Bantu Languages*. Monograph Series 68. Cape Town: The Centre for Advanced Studies of African Society.
- NCDC.** 2003. *The "O" Level Local Language Syllabus. Senior 1–Senior 4*. Kampala: National Curriculum Development Centre.
- NCDC.** 2006. *THEMA. The Newsletter of the Thematic Primary Curriculum. Issue 1. August 2006*. Kampala: National Curriculum Development Centre.
- NCDC.** 2006a. *THEMA News Letter. Issue 2. December 2006*. Kampala: National Curriculum Development Centre.
- Schoenbrun, D.L.** 1997. *The Historical Reconstruction of Great Lakes Bantu Cultural Vocabulary: Etymologies and Distributions*. Sprache und Geschichte in Afrika, Supplement 9. Cologne: Rüdiger Köppe.
- Ssejjengo, E.** 2009. Books: Poetry Ruled. *The New Vision*, 31 December 2009.
- Steehan, S.** 2001. *Kintu: An Annotated Edition of a Lusoga Play*. Unpublished M.A. Dissertation (being a linguistically annotated version of CRC (1998)). Leyden: Leyden University.
- Tendo, S.** 2010. Book Review: Lusoga Gets Dictionary. *Saturday Monitor*, 6 February 2010, p. 22.
- Tucker, A.N. and M.A. Bryan.** 1957. *Linguistic Survey of the Northern Bantu Borderland, Volume 4: Languages of the Eastern Section, Great Lakes to Indian Ocean*. Published for the International African Institute. London: Oxford University Press.
- Tumusiime, D.** 2010. Lusoga Dictionary Printed. *The Observer*, 3 March 2010, p. 23.
- UBS.** 2006. *The 2002 Uganda Population and Housing Census, Analytical Report, Population Composition*. Kampala: Uganda Bureau of Statistics.
- Van der Wal, J.** 2004. *Lusoga Phonology*. Unpublished M.A. Dissertation. Leyden: Leyden University.
- Wambi, M. and R. Naigaga.** 2005. *Idha Tusome* [Come and We Read]. Jinja: Lusoga Language Authority.
- Wandera, S.** 2010. Dictionaries in 3 Local Languages Out. *Daily Monitor*, 22 February 2010, p. 13.
- 湯川, 恭敏** [Yukawa, Yasutoshi]. 2000. ソガ語動詞アクセント試論 [Soga-go dooshi akusento shiron / A Tentative Tonal Analysis of Soga Verbs]. *Journal of Asian and African Studies* 60: 249-290.
- Zgusta, L.** 1971. *Manual of Lexicography*. Janua Linguarum, Series Maior 39. Prague: Academia / The Hague: Mouton.

Addendum 1: "Key to dictionary" in Gonza (2007: vi)

KEY TO DICTIONARY

Abbreviations
RPT = Recent past tense\Participle
n noun
v verb
adj adjective
adv adverb
Lul Lulamooigi
prep preposition
pron pronoun
for Foreign = applied to a word borrowed from another language e.g. : baasi (for) = bus

Finding a word, its meaning and its different forms

Entry ————— **óngérá (okw)** to add, to increase, to give extra; **okwóngérá**
Meaning of entry ————— **émbíró**: to accelerate; **okwóngérá okúsímá**: to
How entry is used ————— deepen; **nhongerá**: add me a bit more; **enhóngézá**:
something extra, an increase; **yóngéráku**: give
something extra; **okwóngéramu**: to add on, to
heighten, to replenish; **okwóngézá**: to cause to
increase, to give extra; **okwóngézááyo**: to postpone

escape v. okwēibá, okúnholoká, okutébúká, okwēsímántúlá,
okúkúúlúlá, okúbomboká, okwēibirirá, okúbombá;
to escape from: okwēnhunhúnkúzá; **to escape with**
force: okúdúúmúúká; **to escape danger/death by**
luck: okúwúnúká, okúwúnúnúká, okúwúnúnsúká
to escape from custody: okwēibá, okútóloká; **n.:**
escape: obútólóse

Addendum 2: Sample from the Lusoga to English side of Gonza (2007: 123)

| | |
|--|--|
| lúfúmó (o) parable, story, fable, metaphor, narrative, legend, myth, folk tale, fiction, pl. enfúmó ; v.: okúfúma : to tell stories | lúgéro (o) story, byword, illustration, legend, maxim, metaphor, myth, proverb, a lie, tale; olúgéro ólwégésa : allegory |
| lúfúngó (o) loincloth | lúgérógéro (o) kind of tree |
| lúfúnhíró (o) a crease | lúgézí (o) sweat, perspiration |
| lúfútúbé (o) multitude of people = éikúúubirá | lúgháánhi (o) strychnia plant used for fencing (Dracoena fragrans, class: Agacaceae) = olúghánó |
| lúgá (o) a cane, club, walking stick = omwígó, whip | lúghaghá (o) wing |
| lúgá (okú) to be unsettled, to move to and fro due to some pressure, to wander around | lúghágála (o) sharpener |
| lugaalúgá (okú) to be unsettled, unstable, to dote on | Lughálálírá the one who is aggressive (nickname) |
| lúgáátá (o) boasting = olúnkulú | lúgháló (o) a task, term of office, turn, shift, public labour = olúghánó ; okúkolá olúgháló : to take your turn; okúkolá mu mpáló : to do it in turns, to do in rotation, to rotate |
| lugabire rubber sandle | lúghangá (o) skull |
| lúgadá (o) reed pl. éngadá | lúgháyó (o) a conversation, chat, discourse = ekíbonó, olúwáyó |
| lúgágáva (o) severe chronic illness, marasmus, severe malnutrition in children, severe | |

Addendum 3: Sample from the English to Lusoga side of Gonza (2007: 325)


| Ff | | English - Lusoga Dictionary |
|--|---|---|
| fable n. olúfúmó, olúgéro | fabricate v. okúkólá ekítú; fabrication: ekikóléibwa | faithfulness: obwésigwá, obwésigé |
| face v. okúlingírírá; in the face: mu máisó; face: máisó; she has a beautiful face: múlúngi mu máisó; to pull one's face: okwénhinkáálá; to pull a wry face: okúdúnkúmálá; to face difficulties: okwétamwá, okwésinwá | fact eky'amazimá; in fact: bwéné, kyoká, mazimá, tí nakíndi; facts: ówinó | fake ekikolereré |
| factional okúyómbágáná, okwesalasalamu | factory n. ékoleró, éigheekézó | falcon wálúgungú |
| fade v. okúsíiwúúká, okúpáálúúká, okúfúmá; to fade away: okúbuláwo, okükendéera, okúwulikiká | faecal m. émpítámbi; watery faeces: olúbyátátá, omúnháláló | fall v. okúgwá; to fall from high up: okúwánúká; to fall sick: okúlwálá; to fall on your back: okúgángálimá; to fall into pieces: okúmáníngúká, okwátikáyátiká, okúmogoká, okúmágáliká; he fell into the water: agwiré mu máadhí dímu; a fall: ekígwó; a heavy fall: omúvúuma; to fall open: okúghendhúká; to fall with a crash: okúwúlúmúká; to fall apart: okúwumbúlúgúká; the baby fell down from being tied secure (on the mother's back): omwána abéekwíké y'agwá ghánsi; to fall by itself/oneself: okwérindimúlá; he/she has fallen on his/her back: agwiré mágáláma |
| fail v. okúgwísá, okúgwá, okúsuulá, okúlémá, okúlémésá, okúfúkúlá; passive: okúlémwá, okúléméérérwá; to fail someone in an exam: okúgwísá; to fail to obtain: okúsubwá; to fail to agree/meet: okúbitambítánwá | fair ekikálámukálámu, kikálámúku; fairly nice: ekikálámukálámu, kikálámúku; pretty | falling-star éibonámulálá |
| | | fallopian tubes endirá |
| | | falls ebyíríró |
| | | false ekífú, ekikolereré, ekítáli kítuufú; falsehood: obúlimbá; false person: omúlimbá; false teeth: ebiinó (of babies between 1 to 5 months allegedly on account of witchcraft) |

Addendum 4: Sample from the front matter in the WSG (Nabirye 2009: 616)

[Note that the front matter was translated into English and included as a back-matter text in the WSG. The section shown here was taken from the back matter.]

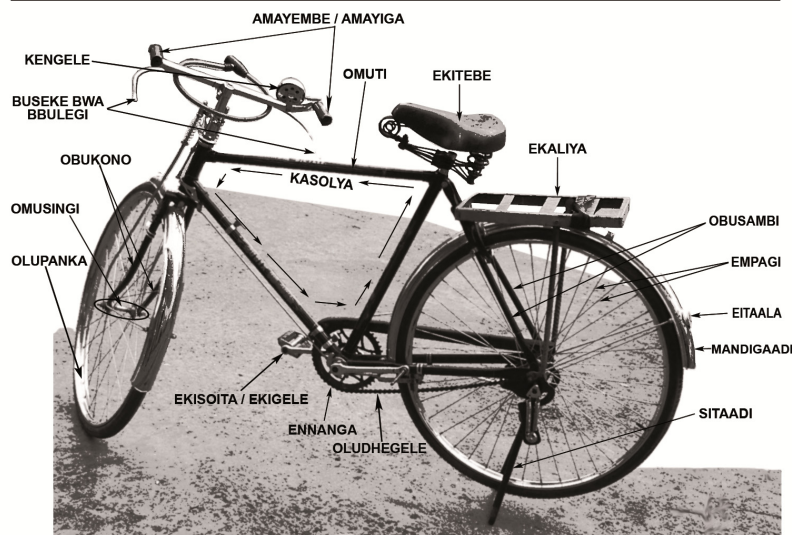
| | | |
|--------------|----------------------|------------------|
| <i>l</i> | liina | “noun” |
| <i>L</i> | ekilanda | “transitive” |
| <i>laz</i> | omulazi | “demonstrative” |
| <i>m</i> | nambeela | “adjective” |
| <i>n</i> | nangeli | “adverb” |
| <i>nhitv</i> | obunhinhitivu | “intensifier” |
| <i>sk</i> | ekisoko | “idiom” |
| <i>T</i> | ekitalanda | “intransitive” |
| <i>t</i> | nantabila | “auxiliary verb” |
| <i>tbk</i> | ekitobeko | “variant” |
| <i>w</i> | akawango | “affix” |
| <i>weile</i> | obuweile | “past tense” |
| <i>wulil</i> | kyakuwulila | “onomatopoeic” |
| <i>y</i> | nakayunzi | “conjunction” |

Addendum 5: Sample from the WSG (Nabirye 2009: 467)

| | |
|--|---|
| <p>(o)lweigulo [(o)lwéigúló] <i>l.</i> [11] bl: [Lg: olweggulo] Ekiseela ky'olunaku eky'eigulo. Bona: NFNN (Olunaku)</p> <p>(o)lwekobaano [(o)lwékóbáánó] <i>l.</i> [11] (o)luyekobaano /-] [11/10] Olukiiko.</p> <p>(o)lweleele [(o)lwéléélé] <i>l.</i> [11(o) luyelele /-] [11] Olukoola.</p> <p>(o)lwema [(o)lwémá] <i>l.</i> [11] (o)luyema /-] Olulembe.</p> <p>(o)lwemba [(o)lwêmbá] tbk: olwêmbó. <i>l.</i> [11] (o)luyemba / 10 (e)nnhembra] bb: [Lsw: wimbo, Lg: oluyimba] Ekya ekiva mu kwemba.</p> <p>(o)lwembo [(o)lwêmbó] <i>l.</i> Bona: (o)lwemba.</p> <p>(o)lwendo [(o)lwéndó] <i>l.</i> [11(o)luyendo/ 10 (e)nnhendo] bl: [Lg: olwendo] 1) Akasumbi akabaaku omukonda nga kamela ku kilime ekilanda. Luba lwa kita-ka nga mulimu ekituli wagati. 2) Akasumbi akasaleku waigulu ab'eila ke baakozesanga okunhwa amaadhi. <i>gez:</i> Mpaayo ku lwendo lw'amaadhi. 3) Ekintu ekikozesebwa okupima amafuta, amata n'ebintu ebindi ebi-</p> | <p>(o)lwisi [(o)lwísí] <i>n.</i> Bona: (o)luusi.</p> <p>(o)lwitamakoli [(o)lwítámakolí] <i>l.</i> [11 / -] Ennhanda ya Kyoga.</p> <p>(o)lwitamalya [(o)lwítámályá] <i>l.</i> [11] (o)luyitamalya /10 (e)ndhitamalya] Enkoofila y'Abasilaamu emyufu ebaaku ekidhefu.</p> <p>(o)lwiwulilo [(o)lwíwúlíló] <i>l.</i> [11] (o)luyiwulilo / 10 (e)ndhiwulilo] bl: [Lg: olujjulilo] Ekifo awali emmele n'endagala nga we baiwula emmele. <i>gez:</i> Ife tuliila ku lwiwulilo.</p>  <p>(o)lwo [(o)lwó] <i>y.</i> [(o)lu- + o] bl: [Lg: olwo] Oluvainhuma oba olunaamala? <i>gez:</i> Buti olwo ?</p> |
|--|---|

Addendum 6: Illustration at the lemma *egaali* 'bicycle' in Nabirye (2009: 45)

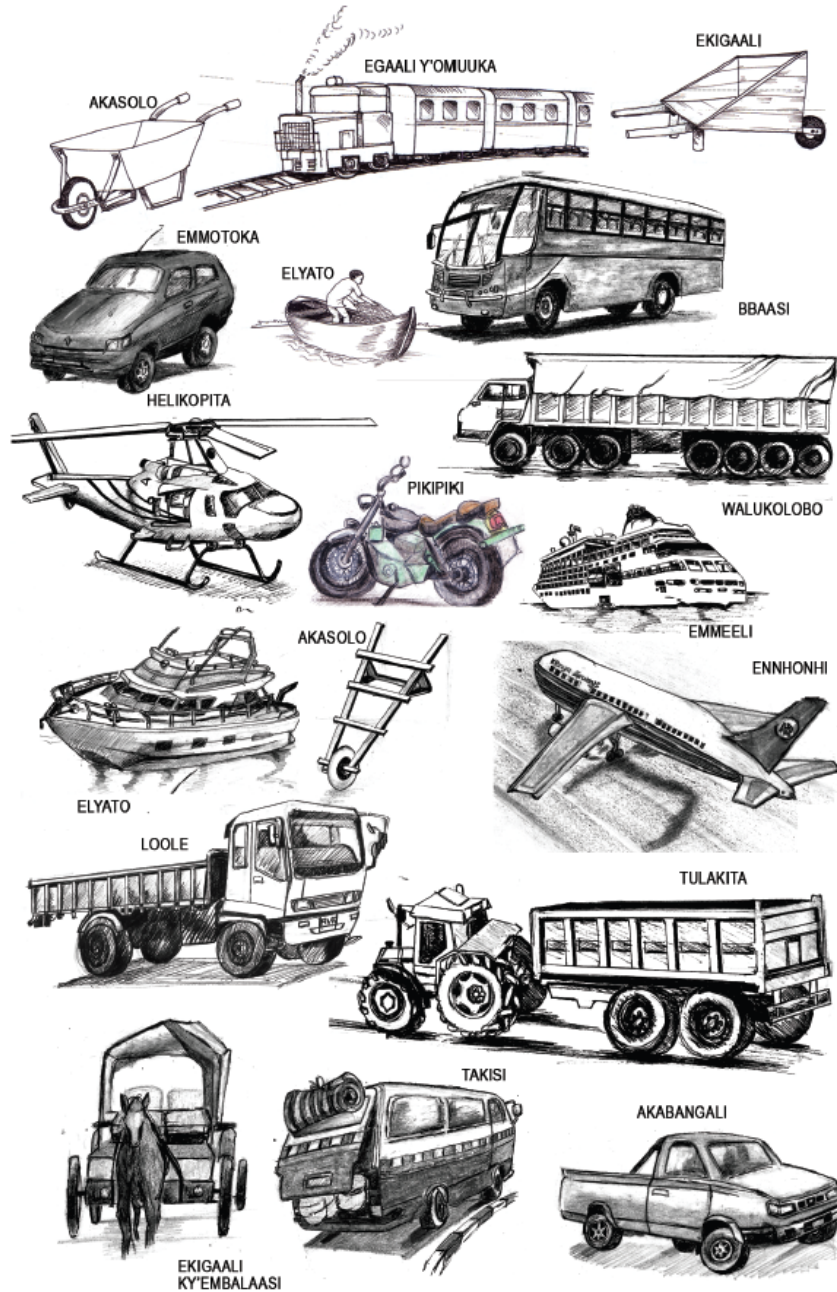
EGAALI



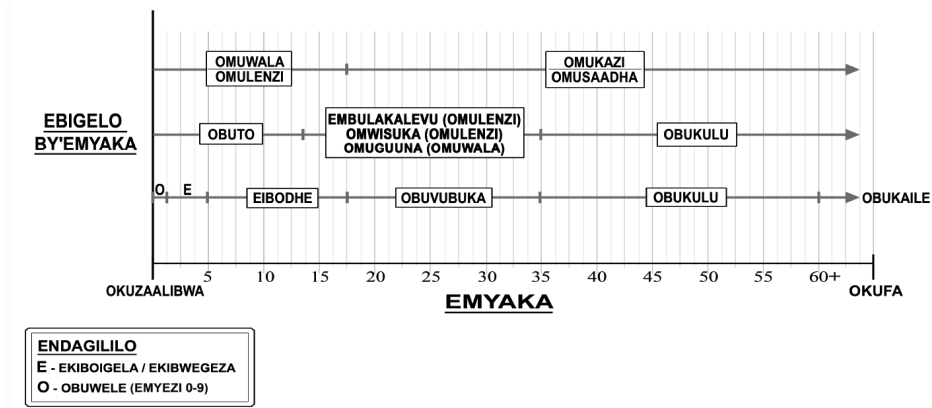
Addendum 7: Transport modes, from the middle matter in Nabirye (2009: D7)

EBYENTAMBULA

D7



Addendum 8: Age ranges, from the middle matter in Nabirye (2009: C1)



Addendum 9: Days in a month, from the middle matter in Nabirye (2009: C2)

| BBALAZA | OLWOKUBILI | OLWOKUSATU | OLWOKUNA | OLWOKUTAANU | OLWOMUKAAGA | SAABBIITI / SAASILA |
|----------------------|--------------------------|---------------------------|-------------------------|----------------------------|----------------------------|-------------------------|
| 1 OLUBELYEBELYE | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 BBALAZA EYABITA | 9 OLWOKUBILI OLWABITA | 10 OLWOKUSATU OLWABITA | 11 OLWOKUNA OLWABITA | 12 OLWOKUTAANU OLWABITA | 13 OLWOMUKAAGA OLWABITA | 14 SAABBIITI EYABITA |
| 15 | 16 EIGULO / IDHO | 17 LEELO | 18 ENKYO / IDHO | 19 DHUZI / LULE | 20 | 21 |
| 22 BBALAZA EIDHA | 23 OLWOKUBILI OLWIDHA | 24 OLWOKUSATU OLWIDHA | 25 OLWOKUNA OLWIDHA | 26 OLWOKUTAANU OLWIDHA | 27 OLWOMUKAAGA OLWIDHA | 28 SAABBIITI EIDHA |
| 29 | 30 | 31 | | | | |

Addendum 10: Day periods, from the middle matter in Nabirye (2009: C2)

| ESAAWA | (7:00 - 11:59) AM | 12:00 | (12:01 - 4:30) PM | (4:31 - 7:00) PM | 7:01 PM - 5:30 AM | (5:31 - 6:59) AM |
|--------------------------|-------------------|-------|-------------------|------------------|-------------------|-------------------------|
| EKIGAMBO KYE' KISEELA | ENKYO | | EMISANA | EIGULO | OBWILE | MADHAMBUIKA / PWIPWIPWI |

Addendum 11: Time periods, from the middle matter in Nabirye (2009: C2)

| | | | | | | | | |
|-------------|----------|-------|---------|----------|--------|--------|--------|--------|
| AKATIKITIKI | DAAKIIKA | SAAWA | OLUNAKU | SABBIITI | OMWEZI | OMWAKA | EKYASA | OMWASA |
|-------------|----------|-------|---------|----------|--------|--------|--------|--------|

The Treatment of Polysemy and Homonymy in Monolingual General-purpose Dictionaries with Special Reference to *Isichazamazwi SesiNdebele*

Eventhough Ndlovu, *Department of Language Management and Language Practice, University of the Free State, Bloemfontein, South Africa*
(*ndlovue@ufs.ac.za / evennthough@yahoo.co.uk*)
and
Sanelisiwe Sayi, *Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe*

Abstract: This article focuses on the treatment of polysemy and homonymy in general-purpose monolingual dictionaries with special reference to *Isichazamazwi SesiNdebele*. It was found that there are some inconsistencies in the treatment of polysemous and homonymous entries in this dictionary. The article shows that an overreliance on one criterion, particularly etymology, to distinguish polysemy and homonymy is often misleading and unreliable. Polysemy itself has its own inherent complexities, among these being the problem of determining the exact number of meanings of a polysemous lemma. When the meanings of a polysemous lemma are listed, the central or primary meaning, which is not always easily ascertainable, should come first. A holistic approach is proposed to distinguish polysemy and homonymy, which entails the use of the following criteria: etymology, relatedness vs unrelatedness of meaning, componential analysis, the identification of the central or core meaning and the test of ambiguity. Whatever results are obtained from a particular criterion, these findings must be compared with those of other criteria, and verified against native speakers' intuitive knowledge and introspective judgements.

Keywords: POLYSEMY, HOMONYMY, METAPHOR, CONCEPTUAL MEANING, ETYMOLOGY, HOMOPHONES, HOMOGRAPHS, LEMMA, SENSE, INTUITION, INTROSPECTION

Opsomming: Die behandeling van polisemie en homonimie in eentalige algemene woordeboeke met spesiale verwysing na *Isichazamazwi SesiNdebele*. Hierdie artikel fokus op die behandeling van polisemie en homonimie in algemene eentalige woordeboeke met spesiale verwysing na *Isichazamazwi SesiNdebele*. Daar is vasgestel dat daar 'n aantal inkonsekwensies in die behandeling van polisemie en homonimie inskrywings in hierdie woordeboek is. Die artikel toon dat 'n te groot steun op een kriterium, veral etimologie, om polisemie en homonimie te onderskei, dikwels misleidend en onbetroubaar is. Polisemie self het sy eie inherente gekompliseerdhede waarvan sommige die probleem is om die presiese aantal betekenisse van 'n polisemie lemma te bepaal. Wanneer die betekenisse van 'n polisemie inskrywing gelys

word, behoort die sentrale of primêre betekenis wat nie altyd maklik bepaalbaar is nie, eerste te kom. 'n Holistiese benadering word voorgestel om polisemie en homonimie te onderskei wat die gebruik van die volgende kriteria behels: etimologie, verwantskap teenoor nieverwantskap van betekenis, die identifikasie van die sentrale of kernbetekenis en die toets van dubbelsinnigheid. Watter resultate van 'n bepaalde kriterium ookal verkry word, hierdie bevindinge moet vergelyk word met daardie van ander kriteria, en uiteindelik geverifieer word met moedertaalsprekers se intuïtiewe kennis en introspektiewe oordeel.

Slutelwoorde: POLISEMIE, HOMONIMIE, METAFOOR, KONSEPTUELE BETEKENIS, ETIMOLOGIE, HOMOFONE, HOMOGRawe, LEMMA, BETEKENIS, INTUÏSIE, INTROSPEKSIE

Introduction

The distinction between polysemy and homonymy is not always clear-cut, therefore remaining a debating-point among linguists and lexicographers. As a result of this debate, a number of criteria have been put forward in an attempt to distinguish these two semantic concepts. However, these criteria have not yielded satisfactory and convincing results. It is against this background that this article proposes a holistic approach in delimiting polysemy and homonymy. It examines the treatment of these two concepts with special reference to the monolingual general-purpose dictionary *Isichazamazwi SesiNdebele* (2001) (henceforth ISN).

As indicated in the front matter (pp. xxii-xxiii), ISN is based on a corpus, so that the words included and the definitions formulated have been done according to the evidence found there. Because the corpus was built at the same time as the dictionary was compiled, in some cases the corpus was still too small to have been useful. This might be the reason why some words and meanings, especially metaphorical meanings, have not been included in ISN. Being a smaller monolingual general-purpose dictionary, the ISN has mostly treated the central or core meanings of words. However, the corpus was effective in deciding on the most commonly used of two or more synonyms.

The editors of ISN mention in the front matter (pp. xxxv-xxxvi) how homonymous and polysemous words are treated. They do not explain, however, on what grounds they distinguished between the two kinds, treating certain entries as homonymous and others as polysemous. The various criteria used to delimit polysemy and homonymy are therefore explored in this article in an effort to examine how dictionary editors, particularly the ISN editors, treated these concepts. In spite of the following discussions in which words are identified as related, it must be remembered that polysemous words may sometimes, for reasons of surveyability or findability, be treated like homonyms in two or more entries.

The article begins with an overview of polysemy and homonymy and then move on to explore the various criteria that may be used to distinguish them. It concludes with findings and recommendations of how lexicographers should treat polysemy and homonymy in monolingual general-purpose dictionaries.

Polysemy

The concept of polysemy involves a number of inherent problems which relate to the difficulty of recognising polysemy, identifying the number of meanings of a polysemous word and dealing with transference of meaning, that is, identifying which is the primary meaning and which is (are) the secondary meaning(s). Linguists and lexicographers seem to agree on what polysemy is. They all define polysemy as a case where the same word has two or more different, but conceptually related meanings or variants of the same meaning (Lyons 1977: 552; Palmer 1981: 101; Hurford and Heasley 1983: 123; Saeed 1997: 64; Zgusta 1971: 61; Jackson 1988: 5; Landau 1984: 100).

A polysemous word has a direct sense from which other senses can, in semantic analysis, be derived by assuming that they are characterised by some added connotation, or by the sense being figurative, or similarly by transference and specialisation (Zgusta 1971: 61). In the case of some highly polysemous words, one of the senses, usually the direct one, is called the dominant. The dominant sense is usually the one which is the first to be thought of by the majority of the speakers of a language if presented with the word out of context (Zgusta 1971: 64).

The meanings attached to a polysemous word are connectable in some way. Such relationships occur in various ways, among others, historically, psychologically and metaphorically (Leech 1974: 228). Two meanings are historically related if they can be traced back to the same source or if the one meaning can be derived from the other (Leech 1974: 228). For example,

- (1) *iganu* (the marula fruit) derives its meaning from the Ndebele word *iganu* (tan). The Nguni people also relate *iganu* (a cow with a tan skin colour) to that resembling the colour of the marula fruit.

These meanings of *iganu*, which are treated as homonyms in ISN are therefore conceptually related in that a marula fruit and a cow of a similar colour are connected with the colour *iganu*. In ISN *iganu* (tan) is not entered as a headword, but is listed under Appendix 4 (p. 543) which forms part of the back matter. It should be the primary sense of the pair in example (1), which are its secondary meanings.

- (2) The three words *phenduka* (repent), *phenduka* (turn around) and *phenduka* (change life) are also treated as homonyms in ISN.

The etymology of *phenduka* (repent) can be traced to the verb *phenduka* (turn around), which is also related to *phenduka* (change life). Through semantic extension, the primary meaning of *phenduka* has come to be used for 'repent' in Christian belief. The meanings of these three verbs are conceptually related because they all share the concept of changing focus or direction. However, in ISN they are treated as homonyms and not as the three meanings of a polyse-

mous word. The primary meaning is carried by *phenduka* (turn around) which should be the main entry, being sense 1 and cross-referenced to its synonym *tshibilika* with the other two meanings listed as senses 2 and 3.

Two or more meanings are psychologically related if present-day users of a language intuitively feel a relationship between them and therefore tend to treat them as different uses of the same word (Leech 1974: 228). The following, for example, are psychologically and historically related:

- (3) *igola* (a wild cat) and *igola* (a nephew), and
- (4) *umthanyelo* (a sweeping broom of grass) and *umthanyelo* (a girl who remains with the bride after the marriage ceremony as her helper, especially for sweeping and other household chores).

In the case of *igola*, the core meaning is that of 'a wild cat' which, through transfer, also came to mean 'a nephew'. A nephew is by descent a child in the uncle's home, who in the patriarchal structure of the Ndebele family, because he is the son of a daughter with a surname from outside, resembles a domestic cat that adopts the traits of a wild cat once it runs away from the home where it was kept. It is a domestic cat, but, like a nephew who cannot claim the inheritance from his uncle, it no longer has legitimate rights. However, in ISN the transferred meaning of *igola* is not entered.

In the case of *umthanyelo*, the core meaning relates to a sweeping broom, while as a helper, the girl assists in sweeping and other household chores. Ndebele speakers perceive that the two meanings are conceptually related, because the girl's chores involve using a broom. This is consequently correctly treated as a case of polysemy in ISN.

The secondary meaning(s) of words historically or psychologically connected is (are) related to or derived from the core or central meaning. It starts with observing striking similarities (though not always) that warrant extension and transference of meaning. Words that are etymologically related are one way or another connected in meaning, either by extension or transfer, among others. According to Zgusta (1971: 76), historically related meanings are however not always psychologically related and neither are psychologically related meanings always historically related. Given this fact, it is therefore necessary to adopt a holistic approach for distinguishing polysemy and homonymy in order to ascertain the findings of one criterion in comparison with those of the other criteria.

The tracing of the various ways in which the senses of a polysemous word relate begins with the speakers' feeling or noting of some striking resemblances between objects or beings as far as features or behaviour are concerned. These striking resemblances then lead to these objects or beings being referred to with the same name. Some sort of similarity is the basis of metaphor, though it is not always cases of visible or striking resemblances that cause metaphoric transfer.

The similarity in a metaphor may be of function, position or many other properties. It does not matter in what direction the first application of a word or the derivation of the literal meaning of a word has taken place.

The word can first apply to humans, animals, plants or objects, e.g. *uhlamvu* (a maize seed) to *uhlamvu* (a bullet) to *uhlamvu* (a phoneme), or *inhloni* (a hedgehog) to *inhloni* (shyness), and transfer can occur in any order, e.g. *iqanda* (an egg) to *iqanda* (a spoiled child), *isambane* (an antbear) to *isambane* (a powerful person) or *inhloko* (an animal head) to *inhloko* (the head of a family) to *inhloko* (the subject of a sentence). This needs to be borne in mind, especially when establishing the literal and transferred meanings. In ISN the transferred meanings of *iqanda* (a spoiled child) and *isambane* (a powerful person) are not included together with the primary meanings of these headwords. The meanings *inhloko* (an animal head) and *inhloko* (the head of a family) are correctly treated as polysemous in ISN, but *inhloko* (the subject of the sentence) which is conceptually related to *inhloko* (an animal head) is entered as homonymous with it.

Lakoff and Johnson (2003: 3) observe that metaphor is pervasive in everyday life, not just in language, but also in thought, and that metaphorical thought is normal and ubiquitous in our mental life, both conscious and unconscious. They further note that fundamentally metaphors are mechanisms of the mind and that our conceptual system is metaphorical in nature.

Metaphor is defined as a case where a word appears to have both a literal and a transferred meaning (Jackson and Amvela 2000: 59-60). Metaphor as a source of polysemy presents inherent challenges in dealing with polysemy itself and in dealing with both polysemy and homonymy. The major challenges comprise identifying the literal meaning of the polysemous word and telling whether it is a case of polysemy or homonymy.

In cases where meanings of polysemous words relate metaphorically, there is a transfer of meaning. In these cases, the polysemous word has both a literal meaning and the transferred or figurative meaning(s) (Jackson and Amvela 2000: 60). According to the speaker's intuitive knowledge or linguistic competence, it is clear which word has the literal sense. The close relationship between the senses of polysemous words can be illustrated by the following examples. A human or animal has *ulimi* (a tongue) and then, because of some resemblance with a certain part of a shoe, the human tongue is metaphorically extended to *ulimi* (the tongue of a shoe). This is given as a qualified entry *ulimi lwesicathulo* in the ISN. In reality it should be the second sense of a polysemous entry *ulimi*. Like *ulimi* (the tongue of a shoe), *ulimi* (a decorative ornament worn by women around their necks) resembles the shape of the tongue. It is a metaphoric extension of the meaning of the tongue. However, in ISN it is treated as a homonymous entry with *ulimi* (a tongue). Since gossiping occurs by using the tongue, it is then also called *ulimi*, the result of a further metaphorical extension. The same is true of the word *isandla* (hand) whose meaning has been transferred to the habit of stealing. A metaphorical extension has occurred here because stealing is done by the hand; hence a thief has *isandla*. Furthermore,

isandla (handwriting) is a metaphorical extension of *isandla* (hand). However, in ISN these transferred meanings are not included.

Metaphor, being the most familiar kind of meaning transfer, is irregular because it applies to individual lexical items, whereas other types of meaning transfer are more regular because they apply to several members of a specific lexical class (Jackson and Amvela 2000: 60). The words for parts of the body provide a good illustration of metaphor, for example, *unyawo* (the human foot) to *unyawo* (the foot of a bed/chair); *ikhala* (the human nose) to *ikhala* (the barrel of a rifle); and *umlomo* (the mouth) to *umlomo* (the opening of a bottle) to *umlomo* (the opening of a boil) to *umlomo* (the habit of provoking others). For the first two examples, the second extended meaning has not been included in ISN. The third example *umlomo* is accurately treated as a polysemous entry. A second homonymous entry *umlomo* is given which, however, is also related to meaning 1 of the first entry.

Owing to striking resemblances in terms of shape, position and many other properties between these objects, meaning transfer took place resulting in cases of polysemy. Each of the senses of these examples are related because the meaning of the other is derived from the one carrying the direct sense. The meanings of the latter words in the pairs have been derived through some discernible process of metaphorical connection.

Except for human body parts, animal and plant parts provide further illustrations of metaphor. For example, *umthala* (the ridge in the paunch of ruminants) to *umthala* (the white layer or line dissecting the sky at night during or towards the rainy season); *ugatsha* (the branch of a tree) to *ugatsha* (a ministry or department); and *uphondo* (an animal horn) to *uphondo* (a cow horn formation) to *uphondo* (a bicycle handle) to *uphondo* (an instrument for injecting traditional medicine to cure *ingubhane* (a type of sickness affecting children)).

Native speakers of Ndebele intuitively know to whom or what these words apply first. As indicated earlier, the direction of meaning transfer does not follow one direction. The literal sense can first apply to humans, animals, plants or objects, for example, *inja* (a dog) to *inja* (a badly behaved person), *intanga* (seeds) to *intanga* (peers), and *isivalo* (a door) to *isivalo* (a student who comes last in class position). In other words, it can be from humans to animals, plants and objects or the other way round. This is often overlooked, leading to the misinterpretation of polysemous as homonymous words.

Metaphor is haphazard, not only within a specific language, but also when the use of the same metaphor is compared across languages. This property renders it irregular (Jackson and Amvela 2000: 60). For example, it may seem obvious that head is appropriate to a nail or eye to a needle in English, but although a nail has a head (*ikhanda*), a needle does not have an eye (*ilihlo*), but a hole (*isikhala*) in Ndebele. A river in Ndebele does not have a mouth (*umlomo*), but a bottle has. These variations show that it is necessary to distinguish polysemy from homonymy in dictionaries. This helps foreign language learners to find particular meanings, since they lack the linguistic competence and

intuitive sense of native speakers to conceive different senses as connected. However, because of their fuzziness and complexity, a distinction between polysemy and homonymy is also necessary for native speakers.

Polysemy involves a number of inherent problems which include recognising it, stating the exact number of meanings of a polysemous word and identifying the core or central meaning of the polysemous word. This is more difficult in cases where the literal meaning is associated with an object and the transferred one with a human being.

- (5) The meaning *inja* (a badly behaved person) is derived by comparison from the basic meaning *inja* (a dog).

In ISN the transferred meaning, which is offensive is not included.

- (6) *Intanga* (peers) and *intanga* (seeds) can serve as an example of a problem in deciding which of the meanings is central.

Many people will question treating *intanga* (peers) as being the transferred meaning. However, it appears *intanga* (seeds) is the central meaning which was transferred to indicate 'people of the same age group', resembling seeds of the same size. In ISN, the plural form *ontanga* (peers) is entered, although the neuter *intanga* is also commonly used. This is a case of figurative extension.

- (7) Regarding *isivalo* (a door) and *isivalo* (a student who comes last in class position), the shared element is the idea of closing. Just as a door closes, so the student who comes last closes the order of merit.

In ISN this transferred meaning is not given, but a second entry has been included with a cross-reference. This was unnecessary, because *isidikiselo* is a synonym of the primary sense of *isivalo* and should have been included directly after the definition of the main entry.

- (8) The words *uphondo* (an animal horn) and *uphondo* (a cow horn formation) are conceptually related. The latter is a transferred meaning of the former because it resembles the half-circle shape of the horns of cows.

In ISN the transferred meanings of *uphondo* (an animal horn) are not included. They both resemble an animal horn in shape: *uphondo* (a bicycle handle) resembles the shape of cow horns, while *uphondo* (an instrument for injecting traditional medicine to cure *ingubhane*) resembles bush buck horns.

These examples reveal the challenges connected with determining polysemy and homonymy, especially if transference of meaning is from objects to humans. The reason is that some native speakers assume that the literal meaning always applies to humans first and the transferred meaning to objects.

Homonymy

Homonymy is defined as a case where two or more words have the same phonological shape and pronunciation, but unrelated meanings (Leech 1974: 228; Lyons 1977: 550; Palmer 1981: 100; Hurford and Heasley 1983: 123; Saeed 1997: 64; Zgusta 1971: 74; Landau 1984: 100; Jackson 1988: 4-5; Svensén 2009: 94). Lexicographers and linguists have often tended to confuse homonyms with homographs and homophones, especially with the former. It is therefore necessary to clearly distinguish homonyms from homographs.

Homographs are words which are spelt the same, but have different pronunciations and unrelated meanings (Saeed 1997: 63; Jackson 1988: 4). For example,

- (9) *ibèlè* (sorghum) and *ibélè* (breast),
- (10) *ímpòndò* (horns) and *ímpóndò* (the old Zimbabwean £1 (*iphawundi*); the Zimbabwean \$2), and
- (11) *bònà* (them) and *bónà* (see).

Homophones are words that are pronounced the same, but have different spellings and unrelated meanings (Saeed 1997: 63; Jackson 1988: 4-5). It appears that homophones such as *sight* and *site*, *meat* and *meet*, *night* and *knight* in English do not exist in Ndebele.

Examples of homonyms, words which are spelt and pronounced the same, but with unrelated meanings, in Ndebele are:

- (12) *impukane* (the house fly) and *impukane* (the shoulder meat of a dead animal which moves during the skinning process),
- (13) *ikhabe* (an ambidextrous person) and *ikhabe* (a watermelon),
- (14) *ifusi* (someone born immediately after twins) and *ifusi* (an abandoned field),
- (15) *umthombo* (malt prepared from sorghum, millet or rapoko for brewing traditional beer) and *umthombo* (a well), and
- (16) *umbala* (colour) and *umbala* (a leg).

In ISN examples (12)–(16) are accurately treated as homonyms. There is no historical connection or psychological relatedness inferable between the senses of these examples. In dictionaries, homonyms are accorded separate headword status, whereas a polysemous entry is treated as a single entry with its definitions listed together and with each definition numbered and descending in order of primary to secondary meaning (Landau 1984: 100). Homonyms are part of the macrostructure of the dictionary and polysemy usually affects the microstructure of the dictionary (Landau 1984: 100; Svensén 2009: 96).

Distinguishing between polysemy and homonymy

In the compilation of a dictionary, there should be consistency and uniformity in the treatment of polysemous and homonymous lemmas. The lexicographer has to apply uniform principles in deciding what should be entered under one and the same lemma and how lexical items having identical base forms are to be treated (Svensén 2009: 94). The lexicographer should not try to make the single senses more sharply distinguished from each other or try to make different senses more closely related than is indicated by the data (Zgusta 1971: 64).

According to Zgusta (1971: 74), homonymy is founded on the way the speakers understand and interpret the meaning or the senses of identical forms. He argues that homonymy begins at the point when the speakers of a language are unable to conceive different senses as connected. When the single senses of a word lose their connection, the word and its meaning can be split into two. Zgusta deplors it that the only way to determine this is to rely on the subjective interpretation of the speakers. This article proposes a holistic approach that seeks to ascertain the speakers' subjective interpretations. This approach involves that, if the speakers' interpretations corroborate with etymology, relatedness of meaning, componential analysis, the identification of the central or core meaning and the test of ambiguity, their interpretation will then be accepted.

According to Zgusta (1971: 75), the speakers' interpretation very much depends, among others, on their level of education. In dealing with polysemy and homonymy in Ndebele, the level of education indeed matters, especially when relatedness or unrelatedness of meaning is empirically proved through componential analysis. However, in this case a thorough knowledge of the peoples' culture, way of life and history is also indispensable. Most of the polysemous words in Ndebele can best be understood in this context, together with a thorough knowledge of the language, especially language use. Zgusta (1971: 78) contends that, when dealing with polysemy and homonymy, lexicographers will do well to verify their own opinions by testing the intersubjective opinions of speakers of the language who have a reasonably representative level of education and command of the language. Lexicographers will also do well to consider the lexicographic tradition of the language, if there is any. Unfortunately for Ndebele, ISN is a pioneering monolingual general-purpose dictionary.

In addition to what Zgusta suggests, it is also necessary for lexicographers to consult historical sources and works on the culture and tradition of the language, as well as people knowledgeable about Ndebele culture, history, language use and way of life. This will go a long way in exposing lexicographers to a variety of interpretations that will help them to ascertain their findings. For a corpus-based dictionary like the ISN, books on the material culture and interviews with such knowledgeable people should have been included in the corpus.

However, the distinction between polysemy and homonymy is not always clear-cut. There are instances where it is difficult to decide whether it is a case of polysemy or homonymy. The distinction between polysemy and homonymy is a contentious issue for linguists and lexicographers. However, a number of criteria have been put forward by both linguists and lexicographers in an attempt to delimit them. Among these criteria are the following: etymology, relatedness of meaning, componential analysis, identifying the central or core meaning, and the test of ambiguity.

Etymology

Etymology comprises a historical approach (Svensén 2009: 96). Two meanings are historically related if they can be traced back to the same source or if the one meaning can be derived from the other (Leech 1974: 228; Lyons 1977: 550; Jackson 1988: 127; Landau 1984: 100; Svensén 2009: 96). According to this criterion, homonyms are known to have developed from what were formally distinct lexemes in earlier stages of a language. If the meanings of an orthographic word can be shown to be derived from a common origin, then this can be treated as a case of polysemy even if the resultant meanings diverge considerably (Jackson 1988: 128). In other words, etymology is based on the notion of historical relatedness (Leech 1974: 228).

Etymology has been relied on by linguists and lexicographers. However, the criterion has a number of limitations which makes it difficult to use as a basis for an argument. Firstly, as Palmer (1981: 102) rightly explicates, the history of a language does not always accurately reflect its present state. History can be misleading, especially in cases where the prime source of information is oral tradition. This is even more problematic in cases where languages with a relatively short tradition of writing are dealt with. However, it is still a difficult task even in languages which have written records dating back hundreds of years. Lyons (1977: 550) notes that in these languages, there are many words about whose historical derivation native and non-native speakers are uncertain. This being the case, etymology proves to be less useful in distinguishing homonymy and polysemy.

Raising similar concerns, Landau (1984: 100) remarks that etymology is an uncertain guide since etymologically disparate words have sometimes evolved associated meanings and historically related words have often developed distinct meanings so that the modern speaker regards them as unrelated, for example, *inyanga* (the moon), *inyanga* (a month), and *inyanga* (a traditional healer). For most Ndebele speakers these are homonyms as reflected in ISN. The same is true of *uhlaka* (someone who carries the traditional healer's equipment), *uhlaka* (splints, sticks used to support the joining of broken joints or bones of people or animals; modern-day plaster), *uhlaka* (a wooden or reed mat used to carry a human corpse for burial) and *uhlaka* (a small hut partition used for drying maize on the roof of a hut); *uhlamvu* (a bullet), *uhlamvu* (a maize

seed) and *uhlamvu* (a phoneme); *intaba* (a mountain) and *intaba* (a pile of stones placed on a grave); *isimbo* (an instrument used by traditional healers to dig out their medicine) and *isimbo* (the payment or token of appreciation given to a traditional healer after a person has been helped or healed); *inyoka* (a snake), *inyoka* (ancestral spirits), *inyoka* (a lazy person) and *inyoka* (an evil person); *uhlanga* (a mark that results from a cut made for specific ritual purposes in someone's flesh), *uhlanga* (a ritualistic way of passing the spirit of witchcraft on to someone by making a cut in the flesh), *uhlanga* (a tribe) and *uhlanga* (the central point of an issue); and *inkonyane* (a newly born calf) and *inkonyane* (knock-knees).

Etymologically the senses of these pairs are related as those of the latter words in the pairs are derived from those of the former words. The senses of each pair are conceptually and psychologically related. When the holistic approach is used as will be further demonstrated in this article, these entries are seen as polysemous. For some, homonymy occurs when the speakers of a language are unable to conceive different senses as connected. Divergence of sense has been seen as the overriding factor in determining homonymy (Zgusta 1971: 74). However, in each of these cases, there is no certain standard for determining whether the divergence is sufficient to warrant separate headword status as homonyms rather than status as one polysemous entry. This then calls for something more than etymology.

Relatedness vs unrelatedness of meaning

One of the major criteria used by linguists and lexicographers for distinguishing homonymy and polysemy is relatedness or unrelatedness of meaning. This criterion seems to correlate with the native speakers' feelings that certain meanings are connected or unconnected. It is based on psychological relatedness where native speakers possess the knowledge and intuition that enable them to conceive different senses as connected or unconnected. This is generally based on the notion of linguistic competence which refers to speakers' implicit, internalised knowledge of the rules of their language.

In applying this criterion to draw the distinction between homonymy and polysemy, linguists and lexicographers ask themselves whether any general remarks about the difference of meaning can be made. They ask questions such as: Are regular types of difference found in the meaning of various words? According to this criterion where the differences are regular and to some degree predictable, a case of polysemy rather than homonymy can be ascertained (Palmer 1981: 103).

One of the most familiar relationships between meanings is that of metaphorical or figurative extension (Lyons 1977: 553). In metaphorical or figurative language, the word will have both a literal meaning and one or more transferred or figurative meanings. Lyons (1977: 553) and Palmer (1981: 103) concur that the most striking set of examples is found with words for parts of the

body. This is shown in the discussion of polysemy and metaphor. Intuitively it is clear enough where the literal sense applies, irrespective of whether the literal sense applies in either direction. Ndebele proverbs reflect that in Ndebele transference of meaning was and still is common from objects to humans.

In cases where metaphor or transference of meaning is involved, speakers, because of their linguistic competence, can, though sometimes with difficulty, discern relatedness of meaning. Generally this criterion is based on the native speakers' judgements. If it is the case that most of them see a metaphorical connection between the different senses of what they take to be the same word, the word in question can then justifiably be marked as a case of polysemy. It is however, necessary to ascertain these findings further by means of other criteria to come to a corroborating conclusion.

Nevertheless, there are several problems associated with this criterion, among them the following: Firstly, relatedness of meaning appears to be a matter of degree. It is uncertain how high or far up the scale words have to be related for them to count as cases of polysemy (Lyons 1977: 552). These degrees of relatedness will in most cases vary from one speaker to another, complicating it even more, since speakers might disagree on similar words.

Secondly, there will always be the problem of pre-theoretical indeterminacy where some native speakers will claim to see a connection between certain words, while others will deny that such relatedness exists (Lyons 1977: 552). In other words, linguistic competence will at times yield conflicting and misleading results, making it difficult to rely on relatedness as a criterion for distinguishing polysemy and homonymy. This again calls for something more than etymology and relatedness of meaning.

Attempts have been made to further explicate the notion of relatedness of meaning in terms of componential analysis of the senses of the lexemes (Lyons 1977: 553). Though componential analysis has its own limitations, the decomposition of the various senses of lexemes does somewhat assist in tracing meaning connections. Componential analysis helps to ascertain the results of etymology, relatedness of meaning and intuition, regardless of the challenges of determining or stating the number of components lexemes must share to count as related in meaning. Once there is a component in common, such lexemes will be considered as related and polysemous, irrespective of it being just one component. Once a commonality has been established, the next step will be to determine relatedness through other criteria. Componential analysis in this proposed approach functions as the empirical test for all the results. For example, it is controversial whether the following words are cases of polysemy or homonymy:

(17) *inyanga* (the moon), *inyanga* (a month) and *inyanga* (a traditional healer).

Inyanga (the moon) carries the central meaning while the other two meanings are derived from it. *Inyanga* (a month) is determined by the cycles of the moon. The moon serves to illuminate and brighten the earth and the traditional healer

also clarifies issues in peoples' lives (Sibanda 1998: 137, 173). In meaning *inyanga* (a traditional healer) is related to the moon because, just like the moon, it has the components *ukuthwasa*¹, *ukufa*² and *ukukhanyisa*³ in common. These shared components between the moon and the traditional healer support the view that this is a case of polysemy. Moreover, the traditional healer's operations are largely determined by the cycles of the moon. On *ngelimnyama*⁴ the traditional healer does not heal, because his/her powers are believed to be connected to the cycles of the moon. The commonalities between the moon and the traditional healer reveal that the latter meaning was derived from the former; hence this counts as a case of polysemy.

The Ndebele people count time, seasons and months according to the cycles of the moon. The appearance of a new moon marks the beginning of a new month. When the moon appears for the first time, the Ndebeles say *inyanga ithwasile* (there is a new moon), when the month begins, the Ndebeles say *ithwasile* (a new month has begun) and when *inyanga* (a traditional healer) completes his/her training, the Ndebeles say *inyanga ithwasile* (the traditional healer has graduated). These shared components therefore confirm that these are polysemous words. In ISN, *inyanga* (the moon), *inyanga* (a month) and *inyanga* (a traditional healer) are treated as homonyms.

The same is the case with

- (18) *inyoka* (snake), *inyoka* (ancestral spirits), *inyoka* (a lazy person) and *inyoka* (an evil person).

These four entries are historically related as the latter derive from the former. In Ndebele culture, ancestral spirits, *amadlozi*, are referred to as *inyoka*, because of their links with snakes. People with different ancestral spirits identify with particular snakes and use their bones for divination purposes and as costume accessories. Snakes such as *inhlathu* (the python), and *indlondlo/inyandezulu* (the green mamba) are treated as symbols of family ancestral spirits so that when the latter is seen in the yard of a house, sorghum, millet or rapoko is scattered over it as a sign of welcome and appreciation for the divine presence of the living dead in the home. Sorghum, millet and rapoko are food of the ancestral spirits. They are used for brewing traditional beer for rituals associated with ancestral worship and bringing home the living dead. Some ancestral spirits use snake fat, especially that of the python, as a cure for various illnesses. When planning the bringing back home ceremony and the appeasement ceremony, the Ndebele people consider the habits of the snakes (Bozongwana 1983; Ndlovu et al. 1995; Sibanda 1998). They are often heard saying:

Lesi yisikhathi esihle sokubuyisa lokuthethela ngoba inyoka zisathule; zisesephansi azikaqansi izihlahla ziye emqongo.

(This is the right time for thanksgiving and appeasement of the ancestral spirits because the snakes are still underground; they have not climbed up the trees.)

Snakes are cold-blooded animals that hibernate underground during winter. In view of this explanation, *inyoka* (snake) and *inyoka* (ancestral spirits) must be considered as cases of polysemy, yet in ISN they are treated as homonyms. Furthermore, *inyoka* (a lazy person) and *inyoka* (an evil person), both meanings not included in ISN, are related to *inyoka* (a snake). The Ndebele people have a saying that shows the relatedness in meaning of *inyoka* (a snake) and *inyoka* (an evil person): *Umunt' ololunya lwenyok' egamul' umunt' ingeyikumudla* (Someone who is evil like a snake that bites someone, yet is not going to eat him/her). The same relationship exists between *inyoka* (a snake) and *inyoka* (a lazy person). When alluding metaphorically to a lazy person, the Ndebele people say: *Yinyoka yomuntu kenelisi khon' ukubamb' umthanyelo* (Someone is a snake, such a lazy person that he/she cannot even hold a grass broom to sweep), indicating that snakes are considered the laziest reptiles.

This information shows that these words should be treated as polysemes, and not as homonyms. As such knowledge is embedded in the cultural values of the Ndebele people, this should be reflected in the treatment of these words. Béjoint (2000: 124) notes that dictionaries are reflections and guardians of the moral and cultural ideologies of the society. Concurring with Béjoint, Svensén (2009: 1) notes that dictionaries are cultural phenomena and products of the culture in which they have originated.

Other examples of lemmas treated as homonymous in the ISN, yet are in fact polysemous are:

- (19) *ilanga* (the sun) and *ilanga* (the day).

The Ndebele people count days according to the cycle of the sun. The rising of the sun marks the beginning of a new day, hence *ilanga* (the day) was derived from *ilanga* (the sun). The creation of a separate entry for *ilanga* (the day) was unnecessary since it is conceptually and semantically related to *ilanga* (the sun). It should have been entered as sense 2 of *ilanga* (the sun), then cross-referenced to its synonym *usuku*.

- (20) *inkanku* (a type of rain bird) and *inkanku* (a cow of a colour resembling that of this type of rain bird).

The latter was historically derived from the former because of the resemblance in colour.

- (21) *isimbo* (an instrument used by traditional healers to dig out their medicine) and *isimbo* (the payment or token of appreciation given to a traditional healer after a person has been helped or healed).

Ndebele speakers feel intuitively that the latter meaning is derived from the former, because the token of appreciation given to the traditional healer relates to the help or healing that comes from the use of this digging instrument.

- (22) *ugatsha* (a branch) and *ugatsha* (a department).

The latter is derived from the former because of the relatedness of meaning and resemblance between the two. A branch is part of a tree and a department is part of a larger unit, for example a faculty. Each being a component of something larger, the pair shares a part-whole relationship.

(23) *isikhundla* (the lair of a hare) and *isikhundla* (a position or post).

As the lair of a hare, *isikhundla*, indicates a place of occupation, possession and power for its inhabitant, so *isikhundla* indicates a post or position of authority and command. Hence these are cases of polysemy.

(24) *ingulube* (pig) and *ingulube* (constellation of three stars that follow each other in a linear form during June and July).

The former, *ingulube* (pig), carries the primary meaning from which the meaning of the latter has been derived: The resemblance lies in the order in which these stars appear, just like wild pigs following one another. When these stars disappear, the one that disappears last is caught up by the other constellation of stars called *izinja* (dogs), hence the Ndebele proverb *Evukela muva ibanjwa yizinja* (The pig that follows last is caught by the dogs). This is derived from the wild pigs' habit of following one behind the other. The one that is last is caught by the dogs. The same applies to the star of this constellation called *ingulube* that disappears last, seemingly caught by the following constellation of stars called *izinja*. This is also a case of polysemy.

(25) *umthala* (the ridge in the paunch of ruminants) and *umthala* (the white layer or line dissecting the sky at night during or towards the rainy season).

Etymologically these two senses are related because the former resembles the latter, acting like the paunch ridge which dissects one of the stomachs of ruminants like the white layer or line dissects the sky; hence they are cases of polysemy, not homonymy as treated in ISN.

ISN also has the entry *umthala* (a type of grass growing along river banks), which is not in any way related to the first two; hence this is indeed, as ISN's treatment shows, a homonym of the two polysemous senses of *umthala*.

(26) *intaba* (mountain) and *intaba* (pile of stones placed on a grave).

The meaning of the latter in this pair was derived from the former, because the piled up stones resemble the shape of a mountain, hence these words are etymologically and conceptually related.

(27) *uhlaka* (someone who carries the traditional healer's equipment), *uhlaka* (splints, sticks used to support the joining of broken joints or bones of people or animals; modern-day plaster), *uhlaka* (a wooden or reed mat used to carry a human corpse for burial) and *uhlaka* (a small hut partition used for drying maize on the roof of a hut).

Etymologically and psychologically the examples given under (27) share the concept of giving support in some way or another and are therefore cases of polysemy.

- (28) *uhlanga* (a mark that results from a cut made for specific ritual purposes in a person's flesh), *uhlanga* (a ritualistic way of passing the spirit of witchcraft on to someone by making a cut in the flesh), *uhlanga* (a tribe) and *uhlanga* (the central point of an issue).

The meaning of *uhlanga* (a ritualistic way of passing the spirit of witchcraft on to someone by making a cut in the flesh) derives from *uhlanga* (a mark that results from a cut made for specific ritual purposes in a person's flesh) because both involve the concept of *ukucaba* (a ritual cut made for a specific traditional purpose). When witches pass on the spirit of witchcraft to their next of kin, they do so by *ukucabela* (cutting the flesh of the next of kin and performing some form of ritual to pass on the spirit). The concept of *ukucaba* shared by these words causes them to be classified as polysemous.

Uhlanga (a tribe) and *uhlanga* (the central point of an issue) are also conceptually related to these two senses of *uhlanga*. Actually the central meaning is *uhlanga* (a tribe). It is related to *uhlanga* (the central point of an issue), both sharing the concept of being the originating source. The relationship between the rituals of cutting one's flesh lies in the fact that they are specifically part of a tribe; they run in the blood, i.e. they are specifically performed on people of the same tribe.

There are, however, two other senses of *uhlanga* which are not related to the four just discussed. They are: *uhlanga* (a honeycomb) and *uhlanga* (a container made of a reed, decorated with beads, worn on the neck and used as an instrument for medicine given to someone having a disease or being possessed to help him/her recover). This is a case of polysemy, because the latter resembles the former in form, especially the bead decoration which looks like a honeycomb. However, in ISN a qualified entry *uhlanga lwenyosi* (a honeycomb) is given. There is no need for this, because in speech and writing Ndebele speakers do not use it in this form. These two senses of *uhlanga* should have been treated in a polysemous entry, not as two homonymous entries, the one furthermore qualified.

- (29) *inkonyane* (a newly born calf) and *inkonyane* (knock-knees).

All newly born calves have knock-knees and people with a similar characteristic are said to have *inkonyane*, being like newly born calves. Etymologically and psychologically the two are related through sharing the same concept. In ISN these are treated as separate entries under variant spellings. Some speakers use *inkonyana* and others say *inkonyane* when referring to a newly born calf and knock-knees. In defining *inkonyane* (a newly born calf), ISN should have included *inkonyane* (knock-knees) as second meaning of a polysemous entry,

indicating the variant form *inkonyana* in square brackets directly after the headword.

However, in ISN all these entries are treated as homonymous words yet the application of the criteria of etymology and relatedness of meaning, aided by componential analysis, reveal otherwise. Such a treatment, confusing polysemy and homonymy, causes the dictionary to be less reflective of the Ndebele culture since it is expected of a dictionary to be a guardian of a peoples' culture. It should be a trusted and respected repository of facts about a language and its culture (Béjoint 2000: 124).

Identifying the central or core meaning

Apart from etymology and relatedness vs unrelatedness of meaning, linguists and lexicographers also identify the core meaning of words in order to ascertain whether they are polysemous or homonymous. This criterion is closely connected with that of relatedness vs unrelatedness of meaning, psychological relatedness and etymology. The criterion of identifying the core or central meaning functions very well in dealing with cases of meaning transfer and metaphor. For example,

- (30) *inhloni* (a hedgehog) and *inhloni* (shyness) are etymologically and semantically related.

The resemblance in character between the hedgehog and shy people caused the Ndebele to see them as related. The core meaning referring to a hedgehog was transferred to the characteristics of shy people behaving like a hedgehog hiding its head and face when it sees people.

The same can be said of

- (31) *usungulo* (a type of harmless snake) and *usungulo* (an instrument used for knitting jerseys and an awl used for making or repairing shoes).

These two words are etymologically related. Owing to the resemblance in the shape and size between this type of snake and the knitting instrument, the latter was also called *usungulo*. This is believed to be the transferred meaning because such instruments were not part of the earlier Ndebele culture. A similar case is

- (32) *uhlamvu* (a bullet), *uhlamvu* (a maize seed) and *uhlamvu* (a phoneme).

The Ndebele people saw bullets for the first time during the early days of colonialism, but they already cultivated maize and other cereals, hence the core meaning relates to a maize or other cereal seed. Because of the resemblance between the two, the Ndebele people then derived *uhlamvu* (a bullet) from *uhlamvu* (a maize seed). Like a maize seed, *uhlamvu* (a phoneme) refers to a single item within the same family unit. These three senses which should have

been treated as a polysemous entry, are now given as three homonymous entries in ISN.

All these examples show how observant the Ndebeles were and still are of the habits of birds, animals, people and nature, as further confirmed by the many proverbs they have. Polysemy provides a look into the culture, history, religion, philosophy and way of life of the speakers of a language. This confirms the idea that humans do most of their speaking in metaphors.

The test of ambiguity

Finally, linguists and lexicographers also use the test of ambiguity to delimit polysemy and homonymy. According to this criterion, homonyms are clearly ambiguous (Palmer 1981: 105). Hurford and Heasley's (1983: 123) definition of homonymy reveals this perception. In defining homonymy, they indicate that it also involves ambiguity:

A case of HOMONYMY is one of an *ambiguous word*, whose different senses are far apart from each other and *not obviously related* to each other in any way.

This definition contains the misconception that homonyms are the only ambiguous words. However, this is not always the case since instances of polysemy yield equally ambiguous sentences. To further support the view that polysemous words also ambiguate sentences, most lexicographers contend that polysemous words should be avoided in definitions since they ambiguate the definitions (Svensén 1993; Landau 1984; Ndlovu 2009). Svensén (1993: 119) notes:

Using near synonyms is entirely valid when the need for semantic precision is not too great. One has to be on guard against synonyms that are polysemous since they ambiguate definitions.

This observation can be illustrated by the following two ISN definitions which Ndlovu (2009: 79) identified as ambiguous because of their use of polysemous words. For example,

- (33) **amachaphaza** [amachaphazi] bz 6. Izibungwana ezimhlophe ezingakabi zinyosi ngamachaphaza.

In this definition, *izibungwana* is polysemous, thus ambiguating the definition. It can be interpreted to mean either the early stages of an insect's development or small insects (in the diminutive sense), which is the primary meaning of *izibungwana*. Apart from being an illustration of a headword that is defined using a polysemous word, *amachaphaza* (bee larvae) and *amachaphaza* (a green, not fully ripe maize cob) are other examples of entries wrongly treated in ISN as homonymous. There are some resemblances between these two: they are both white in colour, and, when being eaten, they both burst (*kuthi patsha pa-*

tsha)⁵, producing a white milky liquid. Because of these resemblances, the secondary meaning of *amachaphaza* (a green, not fully ripe maize cob) was derived from the core meaning of bee larvae.

Another ISN definition that uses a polysemous word is that of *ishinga* (rascal) which is treated as follows:

(34) **ishinga** bz 5. *Ishinga* ngumuntu ongezwayo. FAN *ihlongandlebe*.

The use of the polysemous word *ongezwayo* ambiguates the definition. *Ongezwayo* can be interpreted in two ways, meaning either someone who is naughty or stubborn (*isiqholo*) and/or someone who is deaf (*isacuthe*), which is the primary sense of *ukungezwa*.

When the test of ambiguity is borne in mind, it therefore becomes clear that this cannot be relied on since both polysemy and homonymy can yield ambiguous sentences, as illustrated by the following:

(35) *Inyanga ifile*.

Example (35), which is a case of polysemy can be interpreted to mean that either the last quarter of the moon has disappeared or the traditional healer has passed away.

(36) *UDojiwe ulombala omuhle*.

Example (36), which is a case of homonymy can be interpreted to mean either that Dojiwe has a beautiful complexion or that Dojiwe has gorgeous legs.

Conclusion

In the foregoing discussion, it has been observed that ISN treated some polysemous words as homonymous. An application of the holistic approach to the words identified as homonymous, reveals that they are in fact polysemous. It has further been observed that these criteria indicate that saying words are etymologically related means that speakers have noted relatedness in meaning. By using componential analysis, psychological relatedness of meaning can be empirically proven.

In spite of the fact that polysemy and homonymy are often difficult to differentiate, lexicographers must study them from the point of view not only of "pure" semantics by analysing the lexical meaning of isolated words, but also of the cultural context of the language in question. A deeper understanding of the culture to which the language belongs will yield valuable results. A thorough knowledge of the language, especially as far as idiomatic usage is concerned, is also indispensable for a successful delimitation of polysemy and homonymy. This will guard lexicographers against interpreting some phenomena as indications of polysemy or homonymy while they are not. The inconsistent treatment of polysemy and homonymy in ISN reflects the problematic nature of

attempting to distinguish polysemy and homonymy. It, however, compromises the dictionary's value of being a guardian of a peoples' culture and a trusted and respected repository of facts about a language and its culture.

Endnotes

1. *Ukuthwasa* in reference to the moon indicates that there is a new moon and for the traditional healer it means he/she has completed training as a traditional healer.
2. *Ukufa* in reference to the moon indicates the period when the last quarter of the moon disappears to the time before its reappearance as a new moon and for the traditional healer it means he/she is dead.
3. *Ukukhanyisa* in reference to the moon indicates that it gives light and for the traditional healer it means he/she makes issues clear; he/she discerns.
4. *Elimnyama* refers to the day when the last quarter of the moon disappears. During this day Ndebele people do not wash their bodies after sunset because it is believed that bathing after sunset on this day will cause bad luck for the person who disobeys this belief.
5. *Patsha* is an ideophone of bursting under pressure (as of a ripe fruit when stepped on).

References

- Béjoint, H.** 2000. *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.
- Bozongwana, W.** 1983. *Ndebele Religion and Customs*. Gweru: Mambo Press in Association with the Literature Bureau.
- Hadebe, S. et al. (Eds.)**. 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Hurford, J.R. and B. Heasley.** 1983. *Semantics: A Coursebook*. Cambridge: Cambridge University Press.
- Jackson, H.** 1988. *Words and their Meaning*. London/New York: Longman.
- Jackson, H. and E.Z. Amvela.** 2000. *Words, Meaning and Vocabulary*. London/New York: Continuum.
- Lakoff, G. and M. Johnson.** 2003. *Metaphors we Live by*. Chicago: Chicago University Press.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. New York: Charles Scribner's Sons.
- Leech, G.N.** 1974. *Semantics*. Harmondsworth: Penguin Books.
- Lyons, J.** 1977. *Semantics 2*. Cambridge: Cambridge University Press.
- Ndlovu, E.** 2009. Sense Relations in the Treatment of Meaning in *Isichazamazwi SesiNdebele*. *Lexikos* 19 Supplement: 71-85.
- Ndlovu, T.M. et al.** 1995. *Imikhuba Lamasiko AmaNdebele*. Gweru: Mambo Press
- Palmer, F.R.** 1981. *Semantics*. Cambridge: Cambridge University Press.
- Saeed, J.I.** 1997. *Semantics*. Oxford: Blackwell Publishers.
- Sibanda, J.** 1998. *Isikithi*. Harare: College Press.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Svensén, B.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

Affirming a Role for Specialised Dictionaries in Indigenous African Languages*

Dion Nkomo, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch and Multilingualism Education Project, Centre for Higher Education Development, University of Cape Town, Cape Town, South Africa (deeouf@yahoo.co.uk)*

Abstract: One of the main problems facing speakers and language practitioners of indigenous African languages is the shortage of appropriate dictionaries for a variety of purposes. This lack results in users consulting any available but inappropriate dictionaries. Quite often, users are disappointed because a wrong dictionary does not normally provide the required assistance. Various functions, which the dictionary may serve, are sought in vain from inappropriate dictionaries and other terminological products. Consequently, the potential of lexicography in general and specialised lexicography in particular, remains unrealised owing to a variety of reasons. This article which mainly discusses the specialised dictionary, draws insights from Wiegand's (1984) general theory of lexicography and the theory of lexicographic functions (Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2002, 2008) to affirm the role of specialised dictionaries in indigenous African languages and also to give insights into how such dictionaries may be produced.

Keywords: LEXICOGRAPHY, LSP LEXICOGRAPHY, METALEXICOGRAPHY, DICTIONARY HISTORY, DICTIONARY CRITICISM, DICTIONARY TYPOLOGY, DICTIONARY STRUCTURE, LSP DICTIONARY, AFRICAN LANGUAGES, LEXICOGRAPHIC FUNCTIONS

Opsomming: Die bevestiging van 'n rol vir gespesialiseerde woordeboeke in die inheemse Afrikatale. Een van die hoofprobleme waarvoor sprekers en taalpraktisyns van inheemse Afrikatale te staan kom, is die tekort aan gepaste woordeboeke vir 'n verskeidenheid doeleindes. Hierdie gebrek lei daartoe dat gebruikers enige bekombare maar ongeskikte woordeboeke raadpleeg. Heel dikwels word gebruikers teleurgestel omdat 'n verkeerde woordeboek nie normaalweg die verlangde hulp verskaf nie. Verskillende funksies wat die woordeboek kan aanbied, word tevergeefs in ongeskikte woordeboeke en ander terminologiese produkte gesoek. Gevolglik bly die potensiaal van die leksikografie in die algemeen en die gespesialiseerde leksikografie in die besonder onverwenslik vanweë 'n verskeidenheid redes. Hierdie artikel wat hoofsaaklik die gespesialiseerde leksikografie bespreek, verkry insigte uit Wiegand (1984) se algemene teorie van die leksikografie en die teorie van leksikografiese funksies (Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2002, 2008) om die rol van gespesialiseerde woordeboeke in die inheemse Afrikatale te bevestig, en ook om insigte te verskaf in hoe sulke woordeboeke gemaak kan word.

* This article is a revised version of a paper presented at the Fourteenth International Conference of the African Association for Lexicography, organised by the Xhosa Department, University of the Western Cape, Bellville, South Africa, 6–8 July 2009.

Sleutelwoorde: LEKSIKOGRAFIE, TSD-LEKSIKOGRAFIE, METALEKSIKOGRAFIE, WOORDEBOEKGESKIEDENIS, WOORDEBOEKKRITIEK, WOORDEBOEKTIPOLOGIE, WOORDEBOEKSTRUKTUUR, TSD-WOORDEBOEK, AFRIKATALE, LEKSIKOGRAFIESE FUNKSIES

1. Introduction

Over the last four decades, at least two factors have had a huge impact on practical lexicography on a global level. The first was the use of computer technologies for language data storage and processing which improved the methodological aspects of dictionary-making. Computer technologies led to, among other outcomes, corpus-based lexicography, and made certain lexicographic tasks easier to undertake. The other factor is the contested development of lexicography into a fully-fledged discipline. This resulted from and in more research being conducted on dictionary making methods and dictionary use. Similarly, the ultimate endeavour has been the improvement of lexicographic practice.

Despite such developments on a global level, lexicographic practice in the indigenous African languages continues to face big challenges. These include the level of development of the languages and a young or even a non-existent dictionary culture. Many people in African communities still hold the view that they do not need lexicographic assistance regarding their native languages. This results in dictionaries in indigenous languages failing to be regarded as the utility tools they are meant to be. While this article argues that dictionaries in general have a very important role to play in the development, acquisition and use of indigenous African languages, it mainly focuses on specialised dictionaries (henceforth LSP dictionaries). In the indigenous African languages, the LSP dictionary genre has thus far received the least attention, with lexicographers, linguists and language planners giving more attention to general dictionaries, which are viewed as language standardisation and documentation tools. Thus more focus is on the documentation and preservation of the languages rather than on their use. When it comes to LSP dictionaries, several questions, including the following, are often raised:

- What would you call an atom in a Zulu dictionary?
- How are you going to define terms for specialised concepts in Shona?
- Why would you produce a Physics dictionary in Ndebele when Physics dictionaries are available in the languages of wider communication used in such specialised fields?
- Who needs specialised dictionaries African languages when the teaching of and practice in special subject fields are conducted in English?

The first two questions raise methodological concerns, implying that without effective strategies, LSP lexicography in indigenous African languages is

doomed to fail or result in the production of substandard lexicographic products. To this extent these concerns are genuine. Unfortunately, they are limited in that they are motivated by an English-biased linguistic orientation which purports that African languages are incapable of handling specialised knowledge. The other two questions are more critical, because they query the very essence of lexicographic practice. If no one needs LSP dictionaries and reasons do not exist for their production, then they should not be discussed as they would serve no useful purpose. The status and functional elevation of indigenous African languages is seen as a precondition for the production and relevance of LSP dictionaries and other reference tools. However, that the speakers of indigenous languages do not need LSP dictionaries in their own languages simply because English is the main language used in specialised fields is problematic. It assumes that in the context of English dominance, the second and third language speakers of English do not rely on their native languages to conceptualise, develop and/or communicate specialised knowledge. Drawing insights from the theory of lexicographic functions (Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2002, 2005, 2008, 2010), this article seeks to demonstrate that LSP dictionaries in indigenous African languages can play a very significant role in the teaching and practice of various subject fields in which English is currently the main medium of communication. However, *role* should not be confused with *function* as defined in the theory of lexicographic functions. Instead, in this article, the role of LSP dictionaries is seen as a collective of lexicographic functions, which the dictionaries may serve, thereby affirming their relevance.

2. In defence of lexicography

In order to put our argument into perspective, it is important to underline the fundamental principles of lexicography which constitute its core as a discipline and separate it from other disciplines, particularly linguistics. From the outset, which is traced back to around 4 000 years ago, lexicography has developed as a problem-solving activity (Al-Kasimi 1977, McArthur 1986, Gouws and Prinsloo 2005, Bergenholtz and Tarp 2003, Tarp 2008). This history has earned dictionaries, the most typical lexicographic products, a prestigious regard as utility products (Wiegand 1984, Bergenholtz and Tarp 2003, Tarp 2000, 2008) and containers of knowledge (McArthur 1986). In the course of the history of lexicography until now, not all dictionaries have successfully fulfilled these fundamental principles. Thus the development of metalexicography has had dictionary criticism as one of its constituent parts (Wiegand 1984: 15). However, dictionary criticism has not always been undertaken within a lexicographically motivated model, with Swanepoel (2008: 208-209) arguing that it generally lacks objectivity, validity and reliability, and Béjoint (2000: 113) observing that quite often it amounts to 'dictionary bashing'. Accordingly, most dictionary criticism has failed, according to Swanepoel (2008: 209), to:

(1) assist readers in their decision-making in acquiring the best dictionaries for their usage needs by presenting them with a well-founded analysis of the positive and negative qualities of a dictionary/dictionaries under review, and (2) assist lexicographers in optimizing the functionality of their dictionaries.

In short, dictionary criticism needs to be consistent with the principles of lexicography. It should not be seen as an end in itself but a means towards the production of functional and more user-friendly dictionaries.

Likewise, it is important to recall that lexicography emerged as a practical activity with the objective of producing utility tools rather than developing theories. Dictionaries were first compiled in a pre-theoretical environment. The theoretical component of lexicography, according to Gouws and Prinsloo (2005: 1), can be regarded as "a relative late-comer because lexicography has originally been associated with the practice of dictionary-making". Metalexicography, or theoretical lexicography, has been developed to enable lexicographic practice (cf. Gouws and Prinsloo 2005: 1, Tarp 2008: 11). Such a development has been welcomed, in the hope that it would ensure the production of functional, appropriate and user-friendly dictionaries to keep users abreast of the demands of the information age.

Whereas lexicographic practice has not generally faced direct opposition, it is the establishment of lexicography as a discipline and the development of the relevant theories which have been strongly contested over the years (cf. Sinclair 1984, Atkins and Rundell 2008: 4, Béjoint 2010: 381, Tarp 2010). This has had serious consequences for the development and appreciation of dictionaries as utility products. As a rule rather than an exception, poor dictionaries have been seen as representative of lexicography and its potential. Accordingly, the affirmation of a role for LSP dictionaries in the indigenous African languages should not be viewed in isolation but in a rather broader framework of defending lexicography as a professional and scientific discipline in which practice and theory ought to accompany each other to produce functional and user-friendly dictionaries. In turn, this makes it important to highlight how lexicography has been unwittingly undermined in academic circles through subjective dictionary criticism. A case in point is an article by Moropa (2004) who appraises a parallel corpus as a terminology resource for Xhosa which, however, amounts to a criticism of lexicography rather than a few selected dictionaries which fail to support translation. She begins by quoting from Pinchuck (1977) that no bilingual dictionary is ever perfect, a very simple fact which metalexicography has demonstrated in respect of all dictionary genres. Competent dictionary users would be satisfied by consulting an appropriate and good dictionary, be it monolingual or bilingual. Because no bilingual dictionary is ever perfect, Moropa (2004: 164) then advises:

When attempting to find an equivalent, a translator would do better to consult a good monolingual dictionary in the source language and also a good monolingual dictionary in the target language, if necessary. In many cases the bilingual dictionary does not supply the most suitable translation equivalent for a certain

context, but it does supply a collection of expressions in the target language from which the user must make a choice. The translator may use the equivalents as guidelines at most. The bilingual dictionary also tends to furnish standardised translations that do not correspond to the full lexical range in the two languages and may therefore be incorrect because of shifts of meaning in both languages. A bilingual dictionary should be used only as a last resort.

This quotation contains statements which have negative implications for lexicography. Such views of the limitations of dictionaries in fulfilling specific needs experienced by certain users in specific situations are not peculiar to Moropa (2004) but shared by many linguists and language practitioners. For example, the statements contained in the above quotation are reminiscent of Bowker and Pearson (2002) who raise the following questions:

How often do you have to consult more than one dictionary in order to find the information you are seeking? How often do you find that the information you are seeking is simply not there? How often do you choose a word in the dictionary without really knowing whether it is the right one?

... no single dictionary will ever provide you with all the information you require ...

It cannot be denied that users may find certain dictionaries deficient as far as certain user needs are concerned, but the implication that dictionaries have no potential to address these needs is quite wrong. It is in this respect that lexicographers and metalexicographers need to clarify certain lexicographic issues which are often discussed on the sidelines of the discipline to improve the awareness of lexicographic products and their roles. Issues such as dictionary typology and dictionary functions are not taken into account in the cited criticism of dictionaries. Accordingly, the following statements are made to defend lexicography, but not specifically the dictionaries criticised by Moropa (2004), although it would even emerge that the dictionaries are unfairly criticised.

- Not every interlingual dictionary is a bilingual or multilingual dictionary. Some dictionaries are regarded as bilingual simply because they provide translation equivalents, but translation equivalents do not elevate a dictionary to the status of bilingual dictionaries (cf. Gouws 2004: 268-269, Burkhanov 2004: 17).
- Not every bilingual or multilingual dictionary is a translation dictionary (cf. Burkhanov 2004: 22, Tarp 2002a). To criticise any bilingual dictionary for not providing all the data categories required to aid translation is only justified if its compiler(s) aimed at producing a translation dictionary. However, it may be possible and economic to integrate the translation function in one bilingual dictionary thereby making it polyfunctional.
- While a bilingual dictionary may provide "a collection of expressions in the target language from which the user must make a choice", it may

also provide data categories which guide users in selecting appropriate translation equivalents in various contexts. Collocations and sense discriminators are good examples (Al-Kasimi 1977: 67-75, Mafela 2005: 275-285, Yong and Peng 2007: 204).

- Proscriptive lexicography now bridges prescriptive and descriptive lexicographic approaches (cf. Bergenholtz 2003, Tarp and Gouws 2008) to address linguistic and cultural anisomorphism in bilingual dictionaries.

These points are counter-arguments against the assertion that "a bilingual dictionary should be used *only as a last resort*" (Moropa 2004: 164). This should not be asserted as a rule because it is based on the inadequacy of selected dictionaries regarding the translation of financial texts. It is even unfair to criticise the dictionaries on the basis that they do not support the translation of financial texts, because none of the dictionaries seem to have been compiled specifically to assist users with the translation of financial texts as one of its functions. They are neither dictionaries of business, finance, commerce or economics. For example, in a preface that has been reproduced in the successive reprints of the *English–Xhosa Dictionary*, Fischer et al. (1985) clarify:

The idea for this *English–Xhosa Dictionary* was conceived during many years of studying and teaching Xhosa, being confronted with the lack of an appropriate English–Xhosa dictionary. When the foundation for the dictionary was laid in 1975 it was planned as a small handbook, aimed at the English-speaking student learning Xhosa ...

We started to expand the entries, trying to meet the needs of the English-speaking student to master Xhosa, and to cover the needs of the Xhosa-speaking student confronted with essay and literature in the English curriculum.

This indicates that this dictionary, which happens to be the best available dictionary to use when translating from English to Xhosa, was primarily compiled to support second language learning. To expect general bilingual dictionaries to achieve a fair coverage and treatment of business LSP is, therefore, to ask for too much. It is a biased way of advancing the argument for a parallel corpus as a terminology resource, because it conveniently ignores the fact that the corpus referred to is a specialised one, i.e. consisting of financial texts. The challenges of building such corpora for an African language such as Xhosa, given translation challenges, some of which are reflected in Madiba (2004) in the form of the inconsistent use of particular legal terms in a Venda version of the constitution of South Africa, are also ignored. Some of the challenges could be effectively overcome if good quality translation dictionaries or LSP dictionaries with a translation function were available. For the purpose of the main argument of this article, one finds Moropa's (2004) article useful in demonstrating that African communities do not have appropriate dictionaries for the translation of specialised texts into African languages. Provided with well-conceived translation dictionaries, professional translators would appreciate the value of dic-

tionaries in their practice. Just like parallel corpora (Madiba 2004, Moropa 2004), LSP dictionaries in indigenous African languages need to be seen as potentially useful tools and resources which may solve problems faced in the development and acquisition of specialised languages as well as the translation of specialised texts. Instead of dismissing one type of product to argue for another, it would be more advantageous to argue for the complementary use of different tools and the elimination of their individual shortcomings. What should always be borne in mind are the challenges of developing such tools and, in view of such challenges, theoretical models need to be devised to ensure that user-friendly products are produced. In this article, the focus will only be on LSP lexicography.

3. Challenges for LSP Lexicography in the African Languages

3.1 Language policies

The language policies of most African nations are unfavourable to LSP lexicography as they assign very restricted roles to indigenous languages. For example, in Zimbabwe, English still dominates both the public and private sectors. Indigenous languages are used sparingly as in courts where they are only accommodated with the aid of interpreters or in education prior to the fourth grade. There is a general paucity of resources and tools designed to support the use of the indigenous languages for these limited roles, so much so that the languages are often deemed deficient in communicating specialised knowledge. This means that instead of expanding their functional area, indigenous languages continue losing more ground to English. Perhaps the unavailability of language resources and tools may be seen as one serious handicap in the implementation of South Africa's multilingual national language policy in which indigenous languages are accommodated as official languages. It is unfortunate that such unavailability is often used as the reason for not implementing policies which promote indigenous languages instead of being used to motivate the development of the relevant tools. In the case of lexicography, LSP dictionaries in English continue to be the problem-solving instruments since the problems which have to be solved will usually be encountered in English which is used in those specific situations involving specialised knowledge. There are, however, situations in which LSP dictionaries in indigenous African languages would be useful for at least providing certain information such as equivalents in indigenous languages. It is only when the situations are concretised that the value of LSP dictionaries in African languages may be acknowledged, and this can be done within a metalexicographically-motivated model.

3.2 Theoretical problems

In African countries, there are several projects that are being undertaken on

individual or institutional level to develop the indigenous languages, particularly their terminologies. Most of them lack clear identification and theoretical guidance. However, what clearly emerges from the majority of them is such a strong lexicographic fear that they have painstakingly, though vainly, tried to avoid any association with the term *lexicography*. Rather, the terms *terminology development* and *terminography* are preferred. While they actually develop terminology, usually by collecting and creating terms, the description and treatment of these terms in preparation for dissemination seem to borrow selectively from lexicography. The preferred term for the intended products is *glossaries*, for the sole reason that the products are limited in range and information. Such criteria for distinguishing between glossaries and LSP dictionaries are largely arbitrary. For example, while Bowker and Pearson (2002: 139) criticise dictionaries, it is remarkable that what they regard as the advantages of glossaries (Pearson and Bowker 2002: 138) and data types contained by glossaries (Pearson and Bowker 2002: 160) also apply to LSP dictionaries. In fact, the majority of those practitioners who prefer the production of glossaries over LSP dictionaries, despite the fact that the distinction between them is not clear, seem to opt for a minimalist approach towards the treatment of LSP. Such an approach reflects an evasion of what they regard as tedious lexicographic processes and incomprehensible lexicographic theories (cf. Tarp 2010). Unfortunately, equally minimal is the assistance provided by the end products. This only vindicates those scholars who argue that the distinction between terminology development/terminography on one hand and LSP lexicography on the other is of no practical use, because a large area of confluence exists between them (Bergenholtz and Nielsen 2006: 282-285, Fuertes-Olivera and Arribas-Baño 2008: 8, Tarp 2000: 190). Hartmann (2005: 85) with good reason calls for mutual respect rather than the love-hate relationship which exists between lexicography and other disciplines with which it has contact. This would facilitate the production of user-friendly products, regardless of whether they are called glossaries or LSP dictionaries.

4. A theoretical model for LSP lexicography in African languages

A theoretical model is needed to help lexicographers identify problems which may be solved by LSP dictionaries. The model will also help in the prioritisation, planning and actual compilation of the dictionaries. More importantly, it should also ensure that the produced dictionaries are functional and user-friendly. The model proposed here is a complementary appropriation of Wiegand's (1984) general theory of lexicography and the theory of lexicographic functions developed by (meta)lexicographers at the Aarhus School of Business in Denmark. This is a cautiously conciliatory undertaking, given the less convincing suggestion by Bergenholtz and Tarp (2003) that Wiegand's general theory of lexicography and the theory of lexicographic functions are opposing theories (cf. Nkomo 2008).

4.1 Wiegand's general theory of lexicography

Wiegand's general theory of lexicography is part of metalexicography (the total metadomain of lexicography or dictionary research), alongside the history of lexicography, research on dictionary use and criticism of dictionaries (Wiegand 1984: 15). Over the years, the appraisal of this theory has been undertaken by various scholars (e.g. Smit 1996, 1998, 2002; Gouws 2001; Gouws and Prinsloo 2005) and will not be repeated here. Only a demonstration of how it may provide a general framework for LSP lexicographic practice in African languages is offered.

The four sections of the theory, namely the general section (purposes of dictionaries, relationships to other theories, principles from the history of lexicography), theory of organisation, theory of lexicographic research on language (data collection, data processing, computer assistance) and theory of the lexicographic description of language (dictionary typology and textual theory for lexicographic texts), need to be taken into account in order to facilitate the production of dictionaries in view of an interplay of factors in a particular lexicographic setting. In this case, African communities are regarded as lexicographic scenes within which LSP lexicography in the indigenous languages needs a clear justification based on the functional role which the dictionaries may serve. It is noteworthy that Wiegand's theory has already inspired LSP lexicographic research in African countries such as South Africa (Smit 1996, 1998), Zimbabwe (Nkomo 2008) and Gabon (Ella 2007). The application of some elements of the theory to determine the role of LSP dictionaries in African languages is discussed in the following subsections.

4.1.1 The relationships of lexicography with other theories

This aspect, which falls under what Wiegand (1984: 15) calls the general section, becomes relevant in the consideration of lexicography and its interdisciplinary contacts, also discussed by scholars like Tarp (2000, 2002, 2010) and Hartmann (2005). Theories in fields such as terminology, translation, language teaching and applied linguistics, among others, should be mentioned. So far, LSP lexicography in the African languages has involved a great deal of term-creation as can be illustrated by the compilation of the Shona biomedical terms dictionary, the Shona linguistic and literary terms dictionary, and the Shona and Ndebele music terms dictionary in Zimbabwe, as well as the production of *Isichazi-magama seziBalo Sezikolo* in South Africa. The reason for this is the acknowledged problem of a lack of terminology in the African languages. In view of this, terminological theories also need to be considered by lexicographers who want to compile LSP dictionaries of acceptable linguistic, terminological and lexicographic standards. What for a long time has been regarded as the fundamental differences between terminography and lexicography has been found to be invalid and of no practical use (Bergenholtz and Nielsen 2003:

282-285, Fuertes-Olivera and Arribas-Baño 2008: 8). In fact, LSP dictionaries may prove to be effective standardisation and dissemination tools for terminology. Yet again, translation theories will always come into play given the fact that terminology has to be developed by, among other means, translation and also that LSP dictionaries will need to consider translation as one of their functions. Language teaching theories will have to be considered so that LSP dictionaries may play a practical role in helping users with the acquisition of the LSPs of certain subjects in both English and the indigenous languages.

4.1.2 The history of lexicography

In African language communities, LSP lexicography, just like general lexicography, may be inspired by the fact that lexicography evolved as a problem-solving enterprise in the lexicographic history of specific communities or languages. In Africa, this history is characterised by an acute shortage of particular dictionary genres (with LSP dictionaries being the scarcest), dictionaries whose production was determined by the limited roles of African languages, dictionaries which display certain supremacist attitudes and dictionaries which were compiled without paying attention to the needs and skills of the users (cf. Awak 1990: 10, Busane 1990: 20, Chabata 2007: 280, Gouws 1996: 99, Nkomo 2008: 10). Many socio-economic, political, cultural and religious changes have since taken place. Globalisation has exerted its influence whereby knowledge development, knowledge acquisition and knowledge dissemination necessitate communication which cuts across all kinds of barriers. In such a context, many changes in the lexicographic scenes of Africa are required. These include the production of various dictionary genres for different purposes. The LSP dictionary genre is one important type in this context and, as in the past, it has to deal with the challenges of the day. For example, because of the improved status of African languages and the need to improve it further, in Zimbabwe and South Africa, LSP lexicography has already begun and should continue with the desired expansion of functional space for the indigenous languages.

4.1.3 The theory of organisation

Elaborating on Wiegand's (1994) metalexigraphy, Smit (1996: 105) states that the purpose of the theory of organisation is to determine the basic rules for organising all the areas of lexicographic activities. It starts from general planning to the eventual publication of a dictionary. Thus, the publication of a dictionary is the culmination of a much more comprehensive set of activities, the so-called lexicographic process (Gouws and Prinsloo 2005: 9).

The organisational structure of the South African lexicographic setting is such that dictionary production may follow an ideal lexicographic process with hierarchical agents responsible for the planning and compilation of dictionaries. At the apex is the Pan South African Language Board (PanSALB) which is

responsible for what is called a primary comprehensive lexicographic process. This concerns long-term strategic planning, support, supervision and coordination of lexicographic activities with other stakeholders in lexicographic practice such as the National Lexicography Units (NLUs) and publishers. Part of the activities of the NLUs and the National Language Service (NLS) would constitute a secondary comprehensive process while others would constitute what could be called specific lexicographic processes. Since the NLUs focus on specific languages while the NLS mainly focuses on specialised language, as part of a secondary lexicographic process, their tasks involve planning at institutional level. This entails resource allocation to different lexicographic projects within an institution. In the case of the NLS it also entails the prioritisation of some projects over others in view of the lexicographic needs in a particular language or subject area. Specific lexicographic processes ultimately focus on specific lexicographic projects, their planning and actual execution. They concern dictionary-specific issues.

The conception of such a set-up, in which the government pledged commitment and support, was based on the recognition of the potential of lexicography in addressing various language-related problems and furthering the ideals of the country's multilingual language policy (Gouws 2003: 224-225). It would appear that this could be realised when the country's officially recognised languages were equipped with different types of dictionaries and when these dictionaries were produced according to the prioritisation of the needs of the language speakers. Reflecting the ideals of the theory of organisation, the South African set-up would ensure that lexicographic practice is undertaken in a coordinated way.

In line with the neatly planned hierarchical structure of the South African lexicographic processes, the NLS has created a few products such as the *Multilingual Mathematics Dictionary* to assist primary school learners. However, the South African lexicographic scene has more products of this nature independently created by publishers, notably the *Oxford Mathematics and Science Dictionary* series for schools. Cambridge University Press also published Wababa, Welman and Press's *Isichazi-magama seziBalo Sezikolo* early in 2010. Yet again, various institutions of higher education such as the Stellenbosch University's Unit for IsiXhosa and the University of Cape Town's Centre for Higher Education, among others, are engaged in the compilation of glossaries intended to help students in learning specialised concepts and LSPs of specialised subject areas. There are many such activities which result in products whose compilers normally call them glossaries or learning and teaching resource books, such as Young et al. (2005, 2009) rather than dictionaries. Notably, the purposes for undertaking the creation of such products may be fulfilled effectively if lexicographic guidance is sought to avoid duplication of efforts and to ensure user-friendliness of the products. The lack of the application of lexicographic theories leads to a failure to exploit all the opportunities for providing the target users with optimum assistance. For example, what most of these glossaries, as

they are called, do is to list English terms and provide them with only translation equivalents and at times brief explanations in both English and the indigenous languages. This gives very little help to users, be it for communication or cognitive purposes. More unsatisfactory is that there is very minimal collaboration between the involved individuals and institutions. This poses the danger of the duplication of efforts, and also competition, although this is not necessarily objectionable as long as linguistic communities are provided with appropriate and user-friendly dictionaries.

4.1.4 Theory of the lexicographic description of language

Under this component of metalexigraphy, Wiegand (1984: 17) considers dictionary typology and dictionary structure. These are important elements of a lexicographic theory which marked a departure from a largely linguistic approach to lexicographic practice. For example, it became apparent that instead of dwelling exclusively on the availability of certain linguistic elements in a dictionary, it was also important to consider the accessibility of data and retrievability of information. However, dictionary typology has always been problematic while dictionary structure has become more comprehensible in relation to the reference needs and reference skills of the target users. Therefore, a functional approach, derived from the theory of lexicographic functions, has emerged to be the best course to dictionary typology and dictionary structure (Gouws and Prinsloo 2005: 55). On this account, it needs to be emphasised that it is not merely the conventional structure which makes one dictionary superior to another, or the size which makes a dictionary superior to a glossary, but the ability to exploit structural features so that more information becomes available and accessible.

4.2 The theory of lexicographic functions

The main feature of the theory of lexicographic functions is that it is user-oriented. As literature on this theory and its application is now abundantly available (cf. Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2005, 2008; Gouws 2007), no elaborate exposition of it will be given here. It suffices to reiterate that the theory takes into account the characteristics of specific users, the typical user-situations, the typical problems users experience in such situations, their needs to address such problems and finally, the lexicographic functions of the available and prospective dictionaries regarding the assistance users may find to solve their problems. The functions are divided into cognitive and communication-oriented functions. Cognitive functions are those by means of which the dictionary provides assistance which enhances or improves the user's knowledge, be it general, cultural or linguistic (cf. Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2005, 2008). Communication-related functions address problems

users encounter in communication-related situations such as text production (writing and speaking), text reception (reading and listening) and translation (cf. Bergenholtz and Tarp 1995, 2003; Tarp 2000, 2005, 2008). According to this theory, nothing is taken for granted, be it the inclusion of data or the methods of presentation in a dictionary or even dictionary typology. The insights drawn from the theory of lexicographic functions affirm the importance of LSP dictionaries in the African languages.

4.2.1 Text reception in English

In many African societies, the use of English as the main language in both the public and the private formal sectors is seen as a hindrance to maximum and successful participation of the majority of people who are native speakers of indigenous languages. The concerns are even more serious in education. In addition to a variety of factors, it is true that some learners do not perform well because they are not proficient in the language of instruction. They do not understand the educators and the textbooks they read. Code-switching and code-mixing in classrooms is one indicator that a native language may be the point of departure for some learners in improving general English proficiency and acquiring various LSPs used in the practice and study of specific subjects (Paxton 2009). In view of this, LSP dictionaries may be produced in indigenous languages so that they serve as communication bridges for learners who have very limited command of English for academic and specific purposes. The inclusion of translation equivalents, definitions and/or explanatory notes in their indigenous languages against English LSP elements may solve text reception problems. Most multilingual glossaries being produced for higher education learners in South Africa have this in mind, although the inclusion of data types in them indicates the different degrees to which the role of indigenous languages in English-dominated language-in-education policies is understood.

4.2.2 Text production in indigenous languages

In indigenous African language activism, there are some who believe that in education and other sectors indigenous languages should be elevated to the status of English. Although such a radical approach needs rethinking, it may be noted that its success presupposes a large-scale production of textbooks and other teaching materials. The task would be difficult without LSP dictionaries which take into account written text production as its function. A more practical approach to language planning would rather begin by encouraging the use of indigenous African languages in speakers' own studies, be it literature or linguistics. Compared to the dictionary of music terms which is the first and currently the only LSP dictionary in Ndebele, the production of a dictionary of linguistic and literary terms in Ndebele on such a basis would have had a more

significant impact on the status and functional elevation of Ndebele in Zimbabwe. It has proven difficult for Ndebele to function maximally as a study subject in schools, colleges and universities, because of an acute shortage of Ndebele textbooks and other teaching materials. While the monolingual general dictionary, *Isichazamazwi SesiNdebele* (Hadebe et al. 2001), contributed to the standardisation of the orthography, vocabulary and terminology (Hadebe 2006), which would in turn facilitate text production, its limited scope and depth on various LSPs indicate that there is a vacuum in LSP lexicography to be filled in Ndebele regarding the text production function in the various specialised subject areas.

4.2.3 Translation of texts into the indigenous languages

The dominance of English and the now urgent multilingual agenda in many African societies, as well as globalisation which necessitates intercultural communication have all increased the role of translation, especially from English into the indigenous African languages. The South African constitution, for example, is available in all eleven official languages. This applies to other government documents and documents produced by non-governmental organisations on issues such as human rights and HIV/AIDS which need to be translated into the languages understood by the people. As indicated by Moropa's (2004) article, translators need tools to help them deal with a variety of challenges they encounter within the range of finding the right translation equivalents and their collocations. Contrary to her assertion that dictionaries need to be used as the last resort, they may actually serve as the translator's first source of consultation, and for this to be worthwhile, better dictionaries need to be produced. All that is required are dictionaries compiled with the translation task as one of their functions. If a dictionary of business, commerce or finance was to be produced with isiXhosa as one of the treated languages, it would do well to consider translation of financial texts from English into isiXhosa as its function. Such a dictionary may be more effective than a parallel corpus, given the various possibilities available for lexicographers, including clear meaning discrimination for translation equivalents and the presentation of subject fields in the front matter.

4.2.4 Provision of special, encyclopaedic and cultural information

If LSP dictionaries may solve some of the communication-related problems briefly discussed in the foregoing, the dictionaries would, one way or another, provide special, encyclopaedic and cultural information about their respective subject fields. This does not mean that an LSP dictionary would automatically realise these cognitive functions by providing information relevant to text production, text reception and translation. These dictionaries may, in the mould of

polyfunctional dictionaries, provide such information which may facilitate the realisation of both communication-oriented and cognitive functions, but a very clear prioritisation strategy would have to ensure that certain functions prevail over others in particular LSP dictionaries.

The Zimbabwean LSP dictionaries produced under the auspices of the African Languages Lexical (ALLEX) Project at the African Languages Research Institute (ALRI) provide both specialised and encyclopaedic information about concepts of their respective subject areas. The music terms dictionaries in Ndebele and Shona, respectively *Isichazamazwi SezoMculo* (Nkomo and Moyo 2006) and *Duramazwi reMimhanzi* (Mheta 2005), for example, do not restrict themselves to the traditional lexicographic definitions (Wiegand 1984: 17). Instead, definitions provide information on the term, the concept represented and even pictorial illustrations of concepts such as music instruments and costumes. In this way, users obtain terminological, special and encyclopaedic information. Similarly, it may be instructive that multilingual glossaries produced by institutions of higher learning in South Africa also provide adequate specialised and encyclopaedic information about subject-field concepts since it is their main endeavour to facilitate concept literacy in indigenous languages. Otherwise most of those that are already available have betrayed this goal by simply providing equivalents in the indigenous languages and very brief definitions, in line with some of the principles of terminological definitions which also argue for the univocity of terms (one-concept-one definition). The limitations of traditional and terminological defining principles for pedagogical LSP lexicography have been demonstrated by Carstens (1997) in an article on a multilingual dictionary of Chemistry for institutions of higher education in South Africa and are also discussed in Fuertes-Olivera and Arribas-Baño (2008).

Given that some subject fields are shaped by historical, religious and practical factors which are specific to certain countries or regions, cultural information also needs to be considered in LSP dictionaries. A law dictionary in South Africa would be a good example of an LSP dictionary in which cultural information would have to be considered. The foundations of South African law dictate that a dictionary of law for South African users would be different from a law dictionary in a country with a different legal system, including the European sources of South African law which do not have South African common law as its constituent.

The cultural factor came into play during the production of all Zimbabwean Shona and Ndebele LSP dictionaries. The two music terms dictionaries mentioned earlier, the Shona biomedical terms dictionary (Mpofu et al. 2004) and the Shona dictionary of linguistic and literary terms (Chimhundu and Chabata 2006) all had to be sensitive to the cultural peculiarities of the subject fields in the Zimbabwean context. This included striking a balance in the selection of lemmata and the provision of more encyclopaedic definitions. A careful selection of defining vocabulary was also required, especially in the case of the biomedical terms dictionary where the dilemma with respect to the treatment

of taboo and offensive terms emerged (Mpofu and Mangoya 2005: 129-130). The compilers of the biomedical terms dictionary also had to consider that the young generation of doctors have limited knowledge and vocabulary related to traditional health practices and beliefs (Mpofu and Mangoya 2005: 118). The importance of providing cultural information in LSP dictionaries for African languages has also been underlined by Smit (1998) in her proposed multilingual and multicultural music dictionary for South African music education. All these factors indicate that LSP dictionaries in the African languages have cognitive functions not necessarily limited to assisting users with specialised information, but also with encyclopaedic as well as cultural information.

5. Conclusion

The main objective of this article has been to affirm the role of LSP dictionaries in African languages. As stated in the introduction, this role is constituted by a collective of lexicographic functions that the dictionaries may fulfil, according to the theory of lexicographic functions. The relevance of LSP dictionaries in the African languages is not necessarily derived from their desired status and functional elevation to the level of languages such as English as this may take too long to be realised. Even in the present context of the continued domination of languages such as English as the main languages of practice and teaching in various subject fields, indigenous African languages may serve as auxiliary media. LSP dictionaries in these languages may be useful for functions such as text production, text reception, translation and acquisition of specialised, encyclopaedic and cultural knowledge. However, in the long term, the dictionaries will prove useful in supporting the status and functional elevation of the indigenous languages. It is important that the production of the prospective dictionaries draw guidance from theoretical lexicography so that they are user-friendly and able to satisfy the needs of the users. If LSP dictionaries in the African languages are produced which serve the functions discussed in this article, then disillusioned language practitioners and general users will only realise that what is currently amiss is the availability and quality of the dictionaries and not dictionaries per se. Subsequently, it will emerge that lexicography in general has the potential of addressing most of the cognitive and communication needs confronting African societies in this information age.

Acknowledgement

I am sincerely grateful to Prof. Sven Tarp for his comments during a conversation after the presentation of this paper at the AFRILEX conference, as well as for comments by Dr Emmanuel Chabata and Prof. Rufus Gouws on an earlier version of the article.

References

- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Awak, M.K.** 1990. Historical Background, with Special Reference to Western Africa. Hartmann, R.R.K. (Ed.). 1990: 8-18.
- Béjoint, H.** 2000. *Modern Lexicography*. Oxford: Oxford University Press.
- Béjoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Bergenholtz, H.** 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Bergenholtz, H. and S. Nielsen.** 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.
- Bergenholtz, H. and S. Tarp.** 1995. *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bowker, L. and J. Pearson.** 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York: Routledge.
- Burkhanov, I.** 2004. Requirements for an "Ideal" Bilingual L1-L2 Translation Oriented Dictionary. *Lexikos* 14: 17-32.
- Busane, M.** 1990. Lexicography in Central Africa: The User-Perspective, with Special Reference to Zaire. Hartmann, R.R.K. (Ed.). 1990: 19-35.
- Carstens, A.** 1997. Issues in the Planning of a Multilingual Explanatory Dictionary of Chemistry for South African Students. *Lexikos* 7: 1-24.
- Chabata, E.** 2007. The African Languages Research Institute: A Milestone in the Development of African Languages. *Lexikos* 17: 278-291.
- Chimhundu, H and E. Chabata.** 2006. *Duramazwi reDudziramutauro neUvaranomwe*. Gweru: Mambo Press.
- Ella, E.M.** 2007. A Theoretical Model for a Fang-French-English Specialised Multivolume School Dictionary. Unpublished D.Litt. Dissertation. Stellenbosch: University of Stellenbosch.
- Fischer, A. et al.** 1985. *English-Xhosa Dictionary*. Oxford: Oxford University Press.
- Fuertes-Olivera, P.A and A. Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography. The Representation of Meaning in English and Spanish Business Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H.** 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset in Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J.D. (Ed.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.
- Gouws, R.H.** 2003. Towards the Formulation of a Theoretically Motivated Model for the National Lexicography Units in South Africa. Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts*: 218-246. London/New York: Routledge.
- Gouws, R.H.** 2004. Monolingual and Bilingual Learners' Dictionaries. *Lexikos* 14: 264-274.
- Gouws, R.H.** 2007. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.

- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hadebe, S.** 2006. *The Standardisation of the Ndebele Language through Dictionary-making*. Harare/Oslo: The ALLEX Project.
- Hartmann, R.R.K. (Ed.)**. 1984. *LEX'eter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*. Tübingen: Max Niemeyer.
- Hartmann, R.R.K. (Ed.)**. 1990. *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–25 March 1989*. Exeter: University of Exeter Press.
- Hartmann, R.R.K.** 2005. Lexicography and its Interdisciplinary Contacts with Special Reference to Onomasiology. *Lexikos* 15: 70-89.
- Madiba, M.** 2004. Parallel Corpora as Tools for Developing the Indigenous Languages of South Africa with Special Reference to Venda. *Language Matters* 35(1): 133-147.
- Mafela, M.J.** 2005. Meaning Discrimination in Bilingual Venda Dictionaries. *Lexikos* 15: 276-285.
- McArthur, T.** 1986. *Worlds of Reference: Lexicography, Learning and Language from the Clay Tablets to the Computer*. Cambridge: Cambridge University Press.
- Mheta, G. (Ed.)**. 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Moropa, K.** 2004. A Parallel Corpus as a Terminology Resource for Xhosa: A Study of Strategies Used to Translate Financial Texts. *Language Matters* 35(1): 162-178.
- Mpofu, N. and E. Mangoya.** 2005. The Compilation of the Shona–English Biomedical Dictionary: Problems and Challenges. *Lexikos* 15: 117-131.
- Mpofu, N. et al. (Eds.)** 2004. *Duramazwi reUrapi neUtano*. Gweru: Mambo Press.
- Nkomo, D.** 2008. Towards a Theoretical Framework for LSP Lexicography in Ndebele with Special Reference to a Dictionary of Linguistic and Literary Terms. Unpublished M.Phil. Thesis. Stellenbosch: University of Stellenbosch.
- Nkomo, D. and N. Moyo (Eds.)**. 2006. *Isichamazwi SezoMculo*. Gweru: Mambo Press.
- Paxton, M.I.J.** 2009. It's easy to learn when using your home language but with English you need to start learning the language before you get to the concept: Bilingual Concept Development in an English Medium University in South Africa. *Journal of Multilingual and Multicultural Development* 30(4): 345-359.
- Pinchuck, I.** 1977. *Scientific and Technical Translation*. London: André Deutsch.
- Sinclair, J.M.** 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Ed.). 1984: 3-12.
- Smit, M.** 1996. *Wiegand's Metalexigraphy as a Framework for a Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa*. Unpublished D.Phil. Thesis. Stellenbosch: University of Stellenbosch.
- Smit, M.** 1998. A Multilingual, Multicultural, Explanatory Music Education Dictionary for South Africa — Using Wiegand's Metalexigraphy to Establish its Purposes, Functions and Nature. *Lexikos* 8: 88-97.
- Smit, M.** 2002. The Systematic Development of Wiegand's Metalexigraphy as Reflected in *Kleine Schriften*. *Lexikos* 12: 290-310.
- Swanepoel, P.** 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2002. Basic Elements of a Lexicographic Theory. Emejulu, J.D. (Ed.). 2002. *Éléments de lexicographie gabonaise. Tome II*: 7-20. New York: Jimacs-Hillman.

- Tarp, S.** 2002a. Translation Dictionaries and Bilingual Dictionaries. Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Tarp, S.** 2004. Basic Problems of Learners' Lexicography. *Lexikos* 14: 222-252.
- Tarp, S.** 2005. The Pedagogical Dimension of a Well-conceived Specialised Dictionary. *Iberica. Journal of European Languages for Specific Purposes* 10: 7-21.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S.** 2010. Reflections on the Academic Status of Lexicography. *Lexikos* 20: 450-465.
- Tarp, S. and R.H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Wababa, Z., K. Welman and K. Press (Eds.).** 2010. *Isichazi-magama seziBalo Sezikolo*. Cape Town: Cambridge University Press.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984: 13-30.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam/Philadelphia: John Benjamins.
- Young, D. et al.** 2005. *Understanding Concepts in Mathematics and Science. A Multilingual Learning and Teaching Resource Book in English, Xhosa, Afrikaans and Zulu. Volume 1*. Cape Town: CALLSSA, School of Education.
- Young, D. et al.** 2009. *Understanding Concepts in Mathematics and Science. A Multilingual Learning and Teaching Resource Book in English, Xhosa, Afrikaans and Zulu. Volume 2*. Cape Town: CALLSSA, School of Education.

Die verifiëring, verfyning en toepassing van leksikografiese liniale vir Afrikaans

D.J. Prinsloo, *Departement Afrikatale, Universiteit van Pretoria, Pretoria,
Suid-Afrika (danie.prinsloo@up.ac.za)*

Opsomming: Leksikografiese liniale vir Afrikaans en die Afrikatale is 'n dekadende oud en word algemeen gebruik in die samestelling van woordeboeke. Die samestellers het dit tot dusver nie nodig geag om hierdie liniale te verifieer of te verfyn nie. Kritiek is egter uitgespreek op die samestelling van die Afrikaanse Linaal en dit word in hierdie artikel opgevolg deur 'n poging tot verifiëring van die bestaande linaal asook 'n herberekening of verfyning van die Afrikaanse Linaal. Vir die verifiëring word 'n sogenaamde *stresfaktor* as die basiese benadering gebruik. Dertien liniale word bereken deur middel van 13 subkorpusse van Afrikaans wat doelbewus só saamgestel is dat dit die mees *ongunstige* toestande vir die bestaande linaal skep ten einde te bepaal tot watter mate sodanige liniale afwyk van die Afrikaanse Linaal. Verfyning van die Afrikaanse Linaal word gedoen deur die gemiddelde van vyf liniale te neem wat gebaseer is op dié tipes tekskategorieë wat internasionaal in sogenaamde gebalanseerde en verteenwoordigende korpusontwerpe voorgehou word. Vir dié doel word korpusse saamgestel van koeranttekste, kreatiewe skryfwerk, religieuse tekste en sowel formele as informele taalgebruik. Die gemiddelde waardes van die Afrikaanse Linaal en die verfynde linaal word dan as 'n nuwe sogenaamde 2010 Linaal vir Afrikaans voorgehou. Ten slotte word die onlangs voltooide dele XII en XIII van die *Woordeboek van die Afrikaanse Taal* kortliks met die Afrikaanse Linaal gemeet.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEKE, ALFABETIESE KATEGORIEË, LEKSIKOGRAFIESE LINIAAL, BALANS, KORPUS, KORPUSDATA, OORBEWERKING, ONDERBEWERKING, KORPUSONTWERP

Abstract: The Verification, Refinement and Application of Lexicographic Rulers for Afrikaans. Lexicographic rulers for Afrikaans and the African languages are a decade in existence and are generally used for the compilation of dictionaries. To date the compilers of these rulers did not feel the need to verify or to refine these rulers. Criticism has however been expressed against the ruler for Afrikaans and this is followed up in this article by an effort to verify the existing ruler and a recalculation or refinement of the Afrikaans Ruler. For verification a so-called *stress factor* is used as the basic approach. Thirteen rulers are calculated by means of 13 sub-corpora of Afrikaans purposely compiled to create the most *unfavourable* circumstances for the existing ruler in order to ascertain to what extent these rulers deviate from the Afrikaans Ruler. Refinement of the Afrikaans Ruler is done by calculating the average of five rulers based on those categories that are internationally being used in so-called balanced and representative corpora. For this purpose, corpora are compiled from newspaper texts, creative writing, religious texts and formal as well as informal texts. The average values of the Afrikaans Ruler and the refined ruler are then presented as a new so-called 2010 Ruler for Afrikaans. Finally the recently completed volumes XII and XIII of the *Woordeboek van die Afrikaanse Taal* will briefly be compared with the Afrikaans Ruler.

Keywords: LEXICOGRAPHY, DICTIONARIES, ALPHABETIC CATEGORIES, LEXICOGRAPHIC RULER, BALANCE, CORPUS, CORPUS DATA, OVER-TREATMENT, UNDER-TREATMENT, CORPUS DESIGN

Inleiding

Aan die begin van die nuwe millennium het Prinsloo en De Schryver sogenaamde *leksikografiese liniale* vir die Afrikatale en vir Afrikaans en Engels ontwerp in reaksie op die talle inkonsekwentheid op makrostrukturele vlak wat betref die balans ten opsigte van alfabetiese kategorieë wat hulle in woordeboeke van dié tale teëgekome het. Leksikografiese liniale is vir die eerste keer in 2002 by die Tiende Konferensie van die European Association for Lexicography (EURALEX) in Kopenhagen formeel aan die internasionale gemeenskap bekendgestel (Prinsloo en De Schryver 2002). Die praktiese gebruik van die liniaal vir Afrikaans is uitvoerig gedemonstreer in Prinsloo en De Schryver (2003). Verdere noemenswaardige geleenthede was die gebruik van die liniaal as vertrekpunt vir verdere evaluering van die WAT (De Schryver 2005) en die publikasie van liniale vir al 11 die amptelike landstale van Suid-Afrika (Prinsloo en De Schryver 2005).

Hierdie liniale is sedertdien aan die Suid-Afrikaanse Nasionale Leksikografie-eenhede gedemonstreer en by die opleiding van leksikograwe geïnkorporeer. Leksikografiese liniale word ook deur talle vryskutwoordeboekmakers gebruik en is ook geïntegreer in die woordeboekprogram *Tshwanelex*, 'n gesofistikeerde rekenaarprogram vir die samestelling van woordeboeke.

In die eerste aantal jare sedert 2000 was die aandag beperk tot die uitwerk van die liniale en die gebruik daarvan in die samestelling van nuwe woordeboeke of vir bywerking/hersiening en kritiese analise van bestaande woordeboeke. Geen aandag is gegee aan die moontlike verfyning van die liniale soos wat meer korpusdata deur die jare beskikbaar geword het nie. Die akkuraatheid van die liniale is aanvanklik ook nie bevraagteken nie. Die ontwerpers het dit derhalwe nie nodig geag om die akkuraatheid van die liniale te verifieer of om hulle te probeer verfyn nie. Kritiek is vir die eerste keer uitgespreek, en wel teen die Afrikaanse Liniaal in Botha (2005) in reaksie op De Schryver (2005) se stellings ten opsigte van oor- en onderbewerking in die WAT.

I do not believe that the inclusion of the desk dictionaries in the ruler is warranted, owing to their inherent deficiencies. I therefore have some doubt whether the data resources on which the ruler is based, can be considered as balanced and can give frequency counts that accurately reflect Afrikaans. (Botha 2005: 78)

Hierdie kritiek teen die Afrikaanse Liniaal het daartoe gelei dat die samestelling van die Afrikaanse Liniaal in die afgelope vyf jaar opnuut onder die loep geneem is en die resultaat in hierdie artikel aangebied word.

Die kernvraag wat gevra moet word vir die verfyning van die liniale is of

verandering van die databasis wat vir die berekening van die liniaal gebruik word, sê byvoorbeeld verskillende, onverwante korpuse, verskillende liniale sal oplewer, en as dit die geval is tot watter mate die liniale sal verskil: ingrypend/totaal verskillend of bloot in 'n geringe mate?

Die doel van hierdie artikel is dus om die liniaal vir Afrikaans te verifieer en te verfyn. Verifiëring geskied met behulp van 'n aantal onverwante liniale wat uit diverse korpuse van Afrikaans saamgestel is en die berekening van 'n verfynde liniaal vir Afrikaans geskied op basis van vyf liniale, elk verteenwoordigend van die kategorieë koerantberigte, kreatiewe skryfkuns, religieuse tekste, formele en informele en gesproke taalgebruik. Die seleksie van hierdie kategorieë is gebaseer op die ontwerpe van die *Brown Corpus of Standard American English* (BROWN) en die *Lancaster-Oslo/Bergen Corpus* (LOB), die *Longman/Lancaster English Language Corpus* en die ICE. Ten slotte word die resente dele XII (P) en XIII (Q–R) van die WAT ooreenkomstig die Afrikaanse Liniaal gemeet. Ten einde die Afrikaanse liniale en in besonder die begrippe verwante en onverwante liniale, in perspektief te stel, word die 11 liniale vir die amptelike Suid-Afrikaanse landstale as vertrekpunt geneem. Die eerste stap is om die 11 liniale voor te stel en hulle met mekaar te vergelyk met spesifieke verwysing na die grootste verskille in die omvang van die alfabetiese kategorieë in die onderskeie tale.

Motivering vir en die ontwerp van leksikografiese liniale

Die besluit om leksikografiese liniale te ontwerp, spruit uit waarnemings deur De Schryver en Prinsloo van oënskynde inkonsekwentheid in lemmaseleksie en ongebalanseerde bewerking van lemmas in woordeboeke, veral ten opsigte van oor- of onderbewerking van alfabetiese kategorieë. Tipiese gevalle is dié waar die leksikograaf die samestelling van die woordeboek oorentoesiasties aanpak en dan stoom verloor wat oorbewerking van die eerste paar alfabetiese kategorieë en onderbewerking van die laaste kategorieë tot gevolg het. Vergelyk Kriel (1983) as 'n tipiese voorbeeld in dié verband ten opsigte van die omvang van die bewerkings in die kategorie A versus T in (1).

(1) Pukuntšu

- aka**, *a.ka.* (-ile, -etše), lieg, leuens vertel, jok, onwaarheid spreek (dial. kyk: *aketša*).
- aka**, *a.ka.* inhaak, vashaak, haak, aanhaak, soen, omarm, lieg, liefkoos; *akwa*, gehaak/ingehaak word; *akēla*, haak vir; *akelana*, mekaar liefkoos, vriendskaplik verkeer; *akelwa*, ingehaak word vir; *akiwa*, ingehaak word; *ake*, *ga*, *sa*, nie (in)haak nie; *akē*, mag/moet haak of inhaak; *moaki*, haker; *baaki*, hakers.
- akalala**, *a ka la.la*, sweef, hang oor, oorhang; *akalalēla*, sweef vir/oor; *akalatša*, laat sweef, vlerke oopsprei om te sweef, *akaladitše*, het laat sweef; *se bone nong go- go wa fase ke ga lona*, hoogmoed kom tot 'n val; *akalatšwa*, genoodsaak om te sweef; *akalalwa*; gesweef word; *akalēla*, hang/sweef oor, wydsbeen staan oor; *akalētše*; het gesweef oor; *moakaladi*, persoon wat sweef.
- akama**, *a ka.ma*, verwonder/verbaas wees; *akamela*, inlaat (bemoei) met; *akametša*, (laat) verbaas, verbasing wek, aangaap, toeroep; *akametšwa*, verbaas/aangeaap word, toegeroep word.

akere, 'a kê.'rê, akker.

aketša, a ke.tša, leuen vertel, lieg, jok; *akeditše*, het (gelieg) 'n leuen vertel; *sa aketše*, nie lieg nie.

akga, a.kaga, werp, gooi, slinger, swaai, beweeg; *akgaakga*, heen en weer beweeg (soos branders), slinger, skommel; *akgaakgwa*, heen en weer geslinger word; - *dialta*, arms swaai, met leë hande loop; - *dinao*, voet in die wind slaan; *akgwa*, beweeg/geslinger word; - *akgêga*, skommel, swaai; -*akgêla*, slinger, swaai, werp; *akgêla*, slinger na/vir, tou om die horings gooi, met 'n vangtoug vang, uitkrap, soos kole uit 'n vuur; *akgelwa*, geslinger word, gevang word met 'n tou; - *dikobo*, klere uitpluk.

tsirikana, 'tsi'ri ka.na, klink.

tsirima, 'tsi'ri.ma, klink, lui, uitspuit, vorentoe spring.

tsirimetša, 'tsi'ri me.tša, laat klink, vasbyt, laat lui, styf vasbind.

tsirinya, 'tsi'ri.nya, laat klink, lui.

tširoga, 'tši ro.ga, wakker skrik, senuweeagtig word, opskrik, moedeloos word.

tširogo, 'tši ro.gô, impuls.

tširoša, 'tši ro.ša, wek, skrikmaak.

Die teenoorgestelde (Atkins: mondelinge mededeling, Maart 2004) kom ook voor waar die leksikograaf 'stadig' begin en dan oorentoesiasies raak soos wat die projek vorder of dit kan geskied as gevolg van beleids- of bestuursveranderinge in die samestelling van 'n woordeboek wat oor 'n lang tydperk saamgestel word.

Leksikografiese liniale is ontwerp om die leksikograaf te help om 'n goeie balans te handhaaf ten opsigte van die relatiewe grootte van die alfabetiese kategorieë ten opsigte van die aantal bladsye en die aantal lemmas wat per kategorie bewerk word. Liniale vergestalt daardie *inherente balans* tussen die alfabetiese kategorieë van elke taal en beantwoord die eenvoudige vraag van hoe groot moet kategorie A in die woordeboek wees in verhouding tot kategorie B, kategorie C, ens. In eenvoudige Afrikaans geformuleer, is die vraag bloot "hoe weet die leksikograaf wanneer kategorie A genoegsaam bewerk is en dit tyd is om aan te skuif na kategorie B". Sodanige balans is veral noodsaaklik in gevalle waar 'n voorafbepaalde maksimum aantal bladsye wat die woordeboek mag beslaan deur die uitgewer gestel is. Hierdie aspek van woordeboekmaak is uiters belangrik — vele individuele samestellers en selfs groot woordeboeke het al deur die jare in die slaggetrap van oor- of onderbewerking van sekere alfabetiese kategorieë in hulle woordeboeke. Svensén, alhoewel slegs met verwysing na die bestudering van woordeboeke as basis vir so 'n balans, stel die beginsel nietemin onomwonde:

A decision must also be made as to what fraction of the whole dictionary each initial letter may occupy, so that the size of the finished dictionary can be kept under control during the course of the project. The percentages for each of the various initial letters in a given entry language are fairly constant, and, if such calculations have not already been done by others, it is wise to examine the distribution in a number of dictionaries. (Svensén 1993: 242)

Thorndike (Landau 2001: 360-362) ontwerp 'n sogenaamde bloksistiem vir die distribusie van woordeboekinskrywings ten opsigte van eerste letters. Hy ver-

deel die alfabet in 105 blokke waarin ongeveer dieselfde gewig aan elke blok toegeken word met die doel om 'n ewewigtige verspreiding van leksikale eenhede deur die alfabet te reflekteer. Thorndike se blokstelsel vir die distribusie van woordeboekinskrywings ken byvoorbeeld vier blokke vir E en 13 blokke aan S toe. Vergelyk Tabel 1 wat 'n uittreksel vir E en S uit die Thorndike-sisteam is.

Tabel 1: Die blokke E en S in Thorndike se bloksisteam vir die distribusie van woordeboekinskrywings per eerste letter (Landau 2001: 361)

| | | |
|--------------------|-----------------------|-----------------------|
| ... | S-81 sau–sd | S-88 splo–stas |
| E-29 e–elk | S-82 sea–seo | S-89 stat–stov |
| E-30 ell–en | S-83 sep–shio | S-90 stow–sucg |
| E-31 eo–exb | S-84 ship–sinf | S-91 such–swar |
| E-32 exc–ez | S-85 sing–smd | S-92 swas–sz |
| ... | S-86 sme–sors | ... |
| S-80 s–sat | S-87 sort–spln | |

Landau (2001: 360) merk tereg op:

If one's word list shows that E has as many entries as S, for example, one should suspect that whoever selected the terms for E was far more permissive than the selector for S, and adjust the word list accordingly.

Absolute versus relatiewe liniaalwaardes

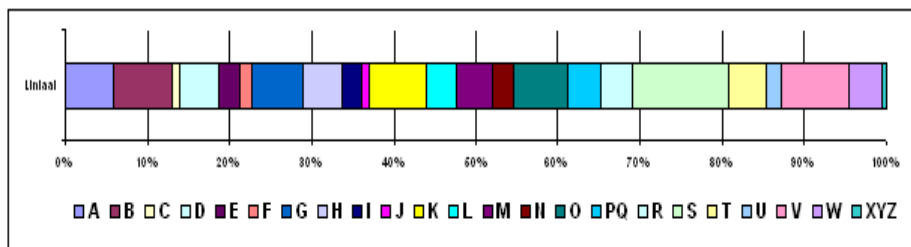
Die doel van Prinsloo en De Schryver met die samestelling van leksikografiese liniale was die daarstelling van 'n meetinstrument vir die grootte van alfabetiese kategorieë as 'n *basiese riglyn*. Alhoewel berekenings per kategorie tot een desimaal afgerond word, suggereer dit nie 'n normeringswaarde dat A 10.0% groot behoort te wees en dat die geringste afwyking verkeerd of ontoelaatbaar is nie. Wat dit wel suggereer, is dat enige substansiële afwyking, sê byvoorbeeld groter as 2%, die moontlikheid van oor- of onderbewerking impliseer en dat dit raadsaam sal wees om dan volgens Landau (2001: 360) die 'woordelys te verstel'. 'n Oorbewerking van 1% van 'n groot kategorie waarvan die liniaal-riglyn 10% is, verteenwoordig 'n absolute oorbewerking van 1% en 'n relatiewe oorbewerking van ongeveer 10%, maar 'n oorbewerking van 1% op 'n liniaal-riglyn van 1% is 'n oorbewerking van 100%. (Vergelyk Prinsloo en De Schryver (2003: 110) vir afsonderlike berekenings van die relatiewe en absolute waardes van die Afrikaanse Liniaal.) As riglyn in die praktiese samestelling van 'n woordeboek is die absolute waarde van meer belang, dit wil sê om te probeer om nie meer as 'n persentasiepunt of twee van die liniaalwaarde af te wyk nie.

Die aanvanklike liniaal vir Afrikaans is gebaseer op twee tipes bronne, woordeboeke en 'n korpus vir Afrikaans, maar gegewe die kritiek van Botha (2005) en die feit dat daar tans baie meer korpusmateriaal beskikbaar is as 'n

dekade gelede, word slegs korpusmateriaal vir die verifiëring en verfyning van die Afrikaanse liniaal in hierdie studie gebruik.

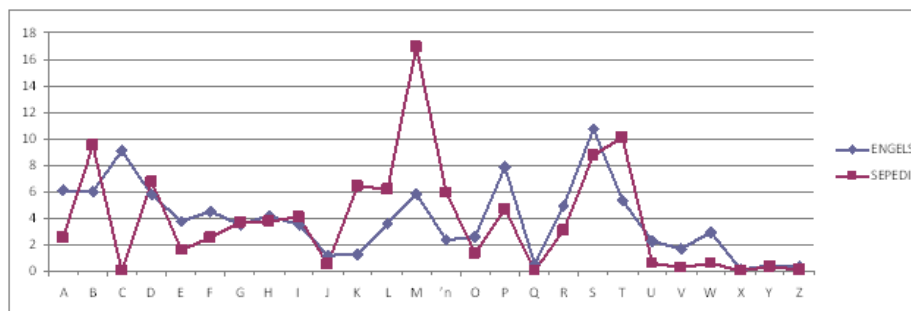
Liniale vir Afrikaans en die ander amptelike landstale van Suid-Afrika

Prinsloo en De Schryver (2005) gee die volledige stel soos in Bylaag A en die liniaal vir Afrikaans soos in Figuur 1. Prinsloo en De Schryver (2002: 488) bevind dat liniaalberekenings op ongelemmatiseerde en gelemmatiseerde korpusdata dieselfde resultate lewer ('n korrelasiekoëffisiëntwaarde $r = 0.991$ (met $r = 1.0$ as die perfekte korrelasiewaarde, nl. twee identiese getallereekse)). Berekenings is gebaseer op ongelemmatiseerde tipes ('types', verskillende woorde) wat in die korpus voorkom en korpusgrootte word ooreenkomstig tekens ('tokens', die aantal woorde in die korpus) aangegee.



Figuur 1: Liniaal vir Afrikaans in % [P en Q; X, Y en Z saamgevoeg] (Prinsloo en De Schryver 2005)

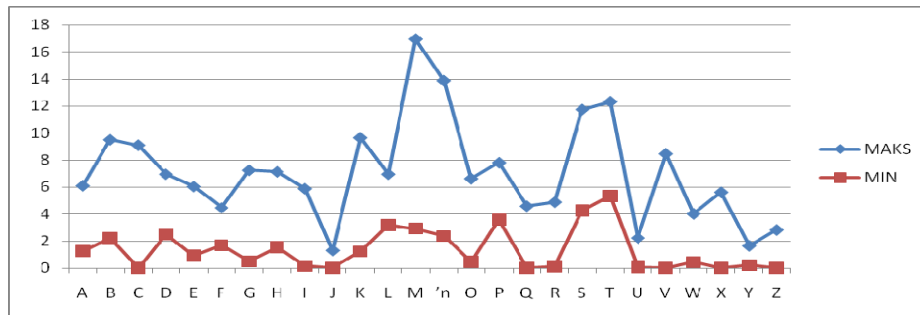
Dit wat 'n gesonde balans vir die grootte van 'n spesifieke alfabetiese kategorie vir taal A verteenwoordig, kan presies die teenoorgestelde vir taal B beteken. So byvoorbeeld is die kategorie **M** in Sepedi baie groot (17%) in vergelyking met Engels (6%), terwyl die kategorie **C** in Engels weer baie groot is (9%) in vergelyking met Sepedi (0%), ens. Vergelyk die liniale vir Engels en Sepedi in Figuur 2 as 'n voorbeeld van *onverwante* liniale.



Figuur 2: 'n Vergelyking van die liniale vir Engels en Sepedi

Die belangrikste basiese beginsel is dat elke taal se liniaal uniek is. Geeneen van die 11 landstale se inherente balans in alfabetiese kategorieë kom ooreen met dié van 'n ander taal nie, selfs nie eers vir nouverwante tale soos Sepedi, Setswana en Sesotho nie (vergelyk Prinsloo 2006).

Die minimum en maksimum waardes per alfabetiese kategorie vir al die landstale word aangedui in Figuur 3.



Figuur 3: Minimum en maksimum liniaalwaardes vir alfabetiese kategorieë van die 11 amptelike landstale

Verifiëring van die Afrikaanse Liniaal

Die blote gemiddelde of die opstapeling van meer liniale wat op willekeurige seleksie van woordeboeke en subkorpuse gebaseer is, sou in beginsel kon bydra tot die verifiëring en selfs tot die verfyning van die Afrikaanse Liniaal. Wat verifiëring betref, is daar besluit om 'n sogenaamde *stresfaktor* as die basiese benadering te gebruik deur korpuse doelbewus só saam te stel dat dit die mees *ongunstige* toestande vir die bestaande Afrikaanse Liniaal skep en dan te bepaal of dit 'n reeks onverwante Afrikaanse liniale tot gevolg het. Vir hierdie doel is 13 subkorpuse saamgestel en 13 afsonderlike liniale bereken. Tabel 2 dui die aard, grootte en samestelling van die subkorpuse aan. 'n Kriptiese omskrywing van hulle negatiewe aard asook hulle korrelasie met die Afrikaanse Liniaal word in Tabel 3 gegee.

Tabel 2: Samestelling van die 13 subkorpuse vir die verifiëring van die Afrikaanse Liniaal

| No. | Aard van / tipe teks | Korpusgrootte in aantal woorde (tekens) | Aantal verskillende woorde (tipes) | Bron |
|-----|----------------------|---|------------------------------------|---|
| 1 | Literêre werk | 105 008 | 10 267 | <i>Versamelde werke</i> C.J. Langenhoven (1933 en 1934) |

| | | | | |
|----|--|-------------|---------|--|
| 2 | Koerantmateriaal: <i>Burger en Beeld</i> , woorde wat meer as een keer voorkom | 141 513 937 | 496 135 | Seleksie uit die <i>Pharos</i> -toets-korpus (<i>Media24</i> -argief) |
| 3 | Koerantmateriaal: <i>Burger en Beeld</i> , woorde wat slegs een keer voorkom | 141 513 937 | 499 551 | Seleksie uit die <i>Pharos</i> -toets-korpus (<i>Media24</i> -argief) |
| 4 | Borduurwerk | 64 968 | 4 609 | <i>Borduursteke vir Suid-Afrika</i> (Eaton 1989) |
| 5 | Laslappie- en appliekwerk | 28 131 | 4 353 | <i>Suid-Afrikaanse boek van laslappie- en appliekwerk</i> (Turpin-Delpport 1988) |
| 6 | Tuinbou | 49 146 | 6 758 | <i>Suid-Afrikaanse tuin</i> (Gilbert 1985) |
| 7 | E-postekste | 4 346 | 1 172 | E-posbus — eie versameling |
| 8 | Religieuse tekste | 919 002 | 14 986 | <i>Afrikaanse Bybel</i> |
| 9 | Pornografie | 10 225 | 2 062 | <i>Loslyf</i> (2000) |
| 10 | Landbou | 8 620 710 | 177 886 | Seleksie uit <i>Landbouweekblad</i> (<i>Media24</i> -argief) |
| 11 | Koerantmateriaal: <i>Rapport</i> | 5 000 829 | 98 290 | Seleksie uit <i>Rapport</i> (<i>Media24</i> -argief) |
| 12 | Akademiese taal | 40 661 | 4 134 | <i>UP Strategiese Plan</i> (2002–2005) |
| 13 | Gesproke taalgebruik | 2 007 | 574 | <i>Kyknet: Robinson Regstreks</i> . April 2010 |

Tabel 3: Die negatiewe aard en korrelasie van die 13 subkorpuse vir die verifiëring van die Afrikaanse Liniaal

| No. | Subliniaalprojeknaam | Negatiewe kriteria | Korrelasiekoëffisiënt |
|-----|----------------------|--|-----------------------|
| | | | $r =$ |
| 1 | Langenhoven | Verouderde Afrikaans, klein korpus | 0.972333 |
| 2 | Pharos > 1 | Domeinspesifiek, hoë(r) frekwensie, onnatuurlike benadering deur die korpus in twee onverwante dele te verdeel op grond van frekwensie | 0.939373 |

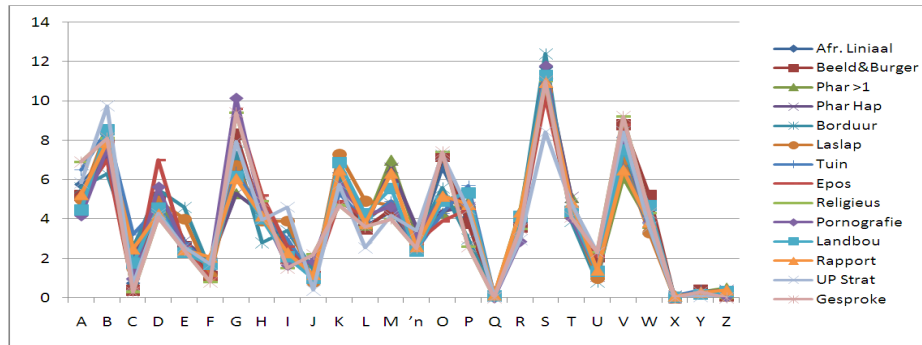
| | | | |
|-----------|-----------------------|--|----------|
| 3 | Pharos = 1 | Domeinspesifiek, baie lae en onnatuurlike frekwensieseleksie | 0.955235 |
| 4 | Borduur | Baie spesifieke onderwerp, herhalende tipiese woordeskat, baie klein korpus | 0.948399 |
| 5 | Laslap | Baie spesifieke onderwerp, herhalende tipiese woordeskat, baie klein korpus | 0.959935 |
| 6 | Tuin | Baie spesifieke onderwerp, herhalende tipiese woordeskat, baie klein korpus | 0.943563 |
| 7 | E-pos | Baie informele teks, baie klein korpus | 0.927709 |
| 8 | Bybel | Domeinspesifiek | 0.947419 |
| 9 | Pornografie | Baie spesifieke onderwerp, herhalende tipiese woordeskat, baie klein korpus, lae register, baie informele teks | 0.942511 |
| 10 | Landbou | Baie domeinspesifiek | 0.968453 |
| 11 | <i>Rapport</i> | Domeinspesifiek | 0.963554 |
| 12 | <i>UP Strat. Plan</i> | Baie formeel, hoë register | 0.928534 |
| 13 | Gesproke taal | Baie klein korpus, gesproke taal, informeel | 0.916724 |
| Gemiddeld | | | 0.947211 |

'n Analise van die verskillende tipes taalgebruik in Tabel 3 dui op besondere verskille. In die geval van die verouderde tekste in 1 is die gebruik van verouderde woordeskat en spelwyses soos *begint* (begin), *had* (het ... gehad), *vammelewe* (vroëer/lank gelede), *posiesie* (posisie), *poliesie* (polisie), *ergens* (êrens), *tamelik* (redelik), *taggentig* (tagtig), *seg* (sê) opmerklik. Ten opsigte van uiteenlopende onderwerpe is dit ook te verwagte dat 'n liniaal slegs gebaseer op laslappiewerk waarskynlik onverwant sal wees aan 'n liniaal saamgestel vir 'n tuinboukorpus of dat die balans in woordeskat tussen die Bybel en 'n pornografietydskrif verskillend sal wees. Die tipiese en herhalende taalgebruik in byvoorbeeld die pornografietekste sentreer rondom 'n klein aantal vulgêre woorde. *Hapax legomena* (woorde wat slegs een keer in 'n korpus voorkom) word in baie studies geïgnoreer as irrelevant vir taalkundige gevolgtrekkings. Korpuslinguïste staan ook skepties ten opsigte van rigtinggewende gevolgtrekkings wat op baie klein korpuse gemaak word soos byvoorbeeld die gesproke-taalkorpus van 2 000 woorde en 'n liniaal wat op slegs 574 verskillende woorde gebaseer is.

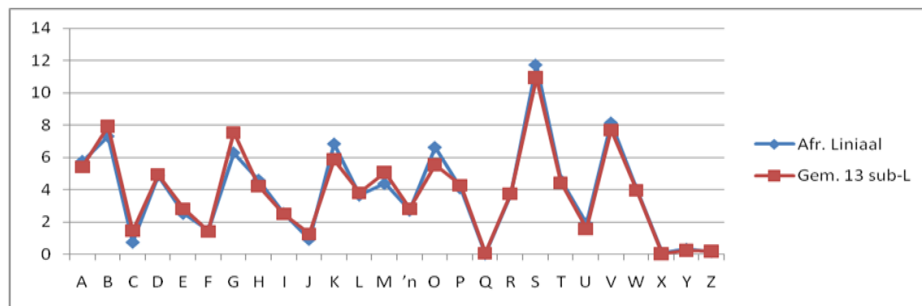
In al die gevalle sou 'n mens verwag dat dié unieke (negatiewe) eienskappe die liniaalwaardes sal versteur.

Dit is egter merkwaardig dat die liniale van elkeen van hierdie 13 uiteenlopende 'onvriendelike' subkorpuse 'n korrelasiekoëffisiënt van hoër as 0.9% met die Afrikaanse Liniaal vertoon en die gemiddelde van hierdie liniale so hoog as 0.95 is soos grafies in Figuur 4 voorgestel word.

Selfs uit hierdie sogenaamde onvriendelike liniale blyk die inherente balans ten opsigte van alfabetiese kategorieë in Afrikaans duidelik en Figuur 5 dui dié noue korrelasie van die Afrikaanse Liniaal met die gemiddelde van die 13 subliniale aan.



Figuur 4: Afrikaanse Liniaal versus 13 onvriendelike Afrikaanse subliniale



Figuur 5: Die Afrikaanse Liniaal versus die gemiddelde van die 13 onvriendelike Afrikaanse subliniale

'n Verfynde liniaal vir Afrikaans

Dit is voor die handliggend dat 'n verfynde liniaal ten beste uit 'n gebalanseerde en verteenwoordigende korpus van Afrikaans ontwikkel moet word, veral in die lig daarvan dat veel meer Afrikaanse teks tans in elektroniese formaat beskikbaar is as 10 jaar gelede.

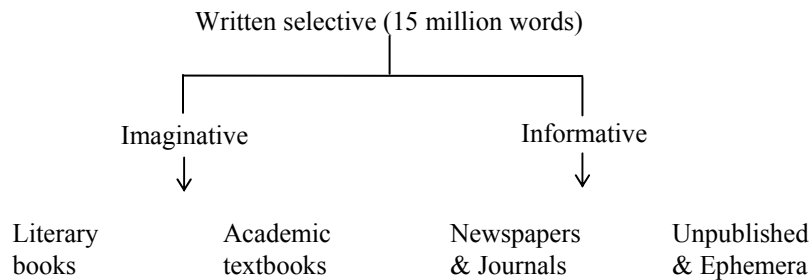
Daar bestaan egter nie 'n waterdigte gebalanseerde of verteenwoordigende korpusontwerp vir enige taal nie. Korpuslinguiste kon tot dusver nie eens eenstemmigheid bereik oor die presiese betekenis in korpusverband van die terme 'gebalanseerd' en 'verteenwoordigend' nie. Bekende korpusontwerpe soos die Brown/LOB-, Longman/Lancaster Oslo Bergen- en die ICE-korpusse in Tabel 4 is bloot pogings om soveel moontlik tipiese taalgebruik in die ontwerp en in die fisiese korpus te inkorporeer. Hierdie debat lê egter buite die bestek van hierdie artikel (sien Biber 1993, Summers 1993, Kilgarrieff 1997, Kennedy 1998, Kruyt en Dutilh 1997, Otlogetswe 2007, en Atkins en Rundell 2008 vir uitvoerige bespreking).

Tabel 4: Brown/LOB, Longman/Lancaster Oslo Bergen en die ICE

(1) Brown/LOB-korpusontwerp

| | |
|-------------------------------|---|
| PRESS: REPORTAGE (44 texts) | FICTION: GENERAL (29 texts) |
| PRESS: EDITORIAL (27 texts) | FICTION: MYSTERY (24 texts) |
| PRESS: REVIEWS (17 texts) | FICTION: SCIENCE (6 texts) |
| RELIGION (17 texts) | FICTION: ADVENTURE (29 texts) |
| SKILLS AND HOBBIES (36 texts) | FICTION: ROMANCE (29 texts) |
| POPULAR LORE (48 texts) | HUMOR (9 texts) |
| BELLES-LETTRES (75 texts) | MISCELLANEOUS: GOVERNMENT & HOUSE ORGANS (30 texts) |
| LEARNED (80 texts) | |

(2) 'n Seleksie van die Longman/Lancaster English Language Corpus



(3) Die ICE-korpusontwerp vir geskrewe tekste

| | | | |
|------------------------------|---------------------|--------------------------------------|---|
| Written Texts (200) | Non-printed (50) | Non-professional writing (20) | untimed student essays (10) student examination scripts (10) |
| | | Correspondence (30) | social letters (15) business letters (15) |
| | Printed (150) | Academic writing (40) | humanities (10) social sciences (10) natural sciences (10) technology (10) |
| | | Non-academic writing (40) | humanities (10) social sciences (10) natural sciences (10) technology (10) |
| | | Reportage (20) | press news reports (20) |
| | | Instructional writing (20) | administrative/regulatory (10) skills/hobbies (10) |
| | | Persuasive writing (10) | press editorials (10) |
| Creative writing (20) | novels/stories (20) | | |

Vir die samestelling van 'n verfynde liniaal word 'n meer simplistiese ontwerp voorgehou wat gebaseer is op die ontwerpe van die Brown/LOB-, Longman/Lancaster Oslo Bergen-, en die ICE-korpusse. Die verfyning geskied op basis van vyf liniale, elk verteenwoordigend van die kategorieë koerantberigte, kreatiewe

skryfkuns, religieuse tekste, formele dokumente en informele en gesproke taalgebruik.

Tabel 5: Korpusontwerp vir die berekening van die verfynde Afrikaanse Liniaal

| Subkorpus | Samestelling | Woorde | Verskillende woorde |
|----------------------------|--|-------------|---------------------|
| Koerantberigte | Koerante: <i>Beeld</i> en <i>Burger</i> | 142 013 488 | 995 686 |
| Kreatiewe skryfkuns | Digbundels, kortverhale | 4 283 294 | 165 599 |
| Religieuse tekste | Preke, Bybeltekste | 975 361 | 16 490 |
| Formele dokumente | Wette, regeringsdokumente | 125 320 | 7 147 |
| Informele en gesproke taal | E-pos, stokperdjies, geselstaal, gesproke taal | 198 225 | 16 950 |

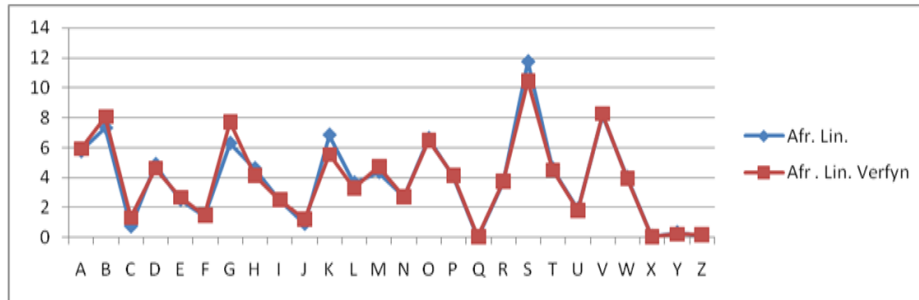
Dieselfde gewig word aan elke kategorie toegeken ongeag die grootte van die subkorpus deur die gemiddeld van die vyf aparte liniale te bereken, vergelyk Tabel 6.

Tabel 6: Die vyf subliniale, die Afrikaanse Liniaal en die Verfynde Liniaal

| | Koerant | Kreatief | Religieus | Formeel | Informeel | Afr. Liniaal | Afr. Lin. Verfyn |
|----------|---------|----------|-----------|---------|-----------|--------------|------------------|
| A | 4.8 | 5.6 | 6.8 | 6.5 | 6 | 5.8 | 5.9 |
| B | 7.5 | 8.3 | 7.9 | 8.7 | 8 | 7.3 | 8.1 |
| C | 2.3 | 1 | 0.3 | 2.1 | 0.9 | 0.7 | 1.3 |
| D | 5.2 | 5 | 4.1 | 4.5 | 4.3 | 4.9 | 4.6 |
| E | 2.7 | 3 | 2.5 | 3 | 2.3 | 2.5 | 2.7 |
| F | 1.9 | 1.6 | 0.8 | 1.6 | 1.6 | 1.5 | 1.5 |
| G | 5.3 | 8.2 | 9.2 | 7 | 8.9 | 6.3 | 7.7 |
| H | 4.3 | 4.1 | 4.8 | 3.4 | 4.1 | 4.6 | 4.2 |
| I | 2.1 | 2.7 | 1.7 | 3.8 | 2.4 | 2.5 | 2.5 |
| J | 1.2 | 1.2 | 2.1 | 0.7 | 0.9 | 0.9 | 1.2 |
| K | 6.5 | 6 | 4.8 | 4.4 | 6 | 6.8 | 5.6 |
| L | 3.7 | 3.3 | 3.6 | 2.5 | 3.4 | 3.7 | 3.3 |
| M | 6.8 | 4.6 | 4.0 | 4.1 | 4.2 | 4.4 | 4.7 |
| N | 3.4 | 2.7 | 2.5 | 3 | 2.1 | 2.7 | 2.7 |
| O | 4.4 | 6.2 | 7.4 | 7.4 | 7.1 | 6.6 | 6.5 |
| P | 4.8 | 4.2 | 2.7 | 4.9 | 4.2 | 4.1 | 4.2 |
| Q | 0.1 | 0.1 | 0.0 | 0.1 | 0 | 0.0 | 0.1 |
| R | 4.2 | 3.5 | 3.4 | 4.2 | 3.3 | 3.7 | 3.7 |

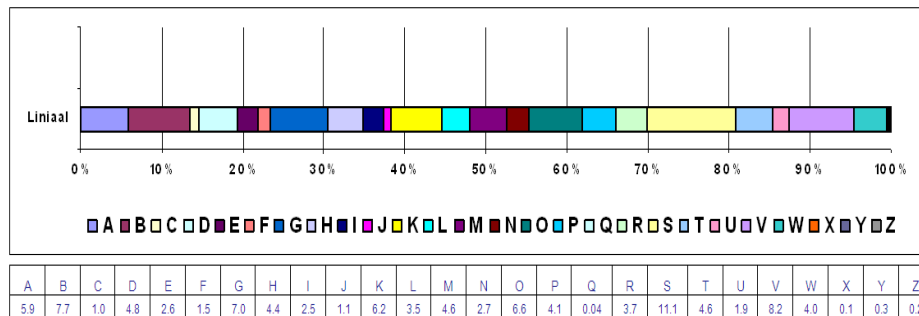
| | | | | | | | |
|---|------|------|------|-----|------|------|------|
| S | 11.4 | 10.1 | 11.0 | 9.1 | 10.6 | 11.7 | 10.5 |
| T | 5.1 | 4.2 | 4.2 | 4.4 | 4.7 | 4.6 | 4.5 |
| U | 1.1 | 1.6 | 2.3 | 2 | 2 | 1.9 | 1.8 |
| V | 6.5 | 8.6 | 9.2 | 8.5 | 8.6 | 8.1 | 8.3 |
| W | 3.9 | 3.9 | 4.4 | 3.7 | 3.8 | 4.0 | 3.9 |
| X | 0.1 | 0 | 0.0 | 0.1 | 0 | 0.1 | 0 |
| Y | 0.3 | 0.3 | 0.3 | 0.1 | 0.2 | 0.3 | 0.2 |
| Z | 0.4 | 0.2 | 0.0 | 0.1 | 0.2 | 0.2 | 0.2 |

Die korrelasiekoëffisiëntwaarde van die Afrikaanse Liniaal en die verfynde liniaal is $r = 0.98$ en dié noue verwantskap word in Figuur 6 grafies aangedui.



Figuur 6: 'n Verfynde liniaal vir Afrikaans

Die gemiddelde van die Afrikaanse Liniaal en die Verfynde Liniaal word in Figuur 7 as die Afrikaanse Liniaal 2010 voorgedui.



Figuur 7: Die saamgestelde 2010-Liniaal vir Afrikaans

Leksikografiese liniale kan benewens 'n persentasiewaarde per letter van die alfabet, ook vir praktiese doeleindes in enige aantal dele uitgedruk word. Prinsloo en De Schryver (2003: 123) verdeel die Afrikaanse liniaal in 179 blok-

ke. In Tabel 7 word die liniaal vir formele dokumente (ook Tabel 6 kolom 5) in 100 dele opgebreek wat elk dus een persent van die liniaal verteenwoordig.

Tabel 7: 'n Bloksisteam vir Afrikaans bestaande uit 100 dele

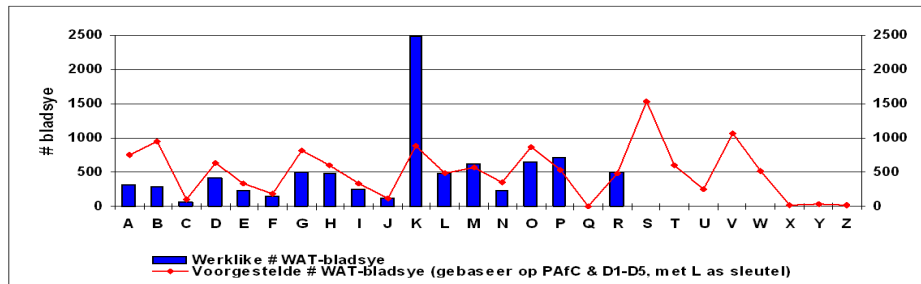
| | | | | | | | | | |
|----|------|----|------|----|------|----|------|-----|------|
| 1 | AANR | 21 | DOKU | 41 | JARE | 61 | OPLE | 81 | SYFE |
| 2 | ABSO | 22 | EENH | 42 | KATA | 62 | ORDE | 82 | TEKO |
| 3 | AFGE | 23 | EKSP | 43 | KLER | 63 | PAPI | 83 | TEWE |
| 4 | AKKR | 24 | ERKE | 44 | KONT | 64 | PERM | 84 | TOER |
| 5 | ANDE | 25 | FAKS | 45 | KUND | 65 | POGI | 85 | TRUS |
| 6 | ARTS | 26 | FORM | 46 | LAND | 66 | PRIM | 86 | UITG |
| 7 | BEDO | 27 | GEBO | 47 | LENI | 67 | PROM | 87 | UITV |
| 8 | BEHO | 28 | GEHA | 48 | LOSI | 68 | RADI | 88 | VARI |
| 9 | BENA | 29 | GEMA | 49 | MANL | 69 | REFE | 89 | VERB |
| 10 | BESI | 30 | GEPA | 50 | MEGA | 70 | REGU | 90 | VERK |
| 11 | BEST | 31 | GESO | 51 | MINI | 71 | RESO | 91 | VERS |
| 12 | BEWA | 32 | GEWE | 52 | MPUM | 72 | RUST | 92 | VERT |
| 13 | BLYW | 33 | GRON | 53 | NASP | 73 | SATE | 93 | VISS |
| 14 | BREE | 34 | HAND | 54 | NETW | 74 | SELF | 94 | VOLW |
| 15 | BYLA | 35 | HERO | 55 | NOTA | 75 | SIMU | 95 | VOOR |
| 16 | CHOR | 36 | HOOF | 56 | OMGE | 76 | SKUL | 96 | VYFD |
| 17 | CV | 37 | IDEN | 57 | ONDE | 77 | SONS | 97 | WATE |
| 18 | DEBA | 38 | INFR | 58 | ONTE | 78 | STAA | 98 | WERK |
| 19 | DESE | 39 | INRI | 59 | OORB | 79 | STER | 99 | WILD |
| 20 | DINO | 40 | INTE | 60 | OORV | 80 | STUD | 100 | ZIMB |

Die bloksisteam kan met vrug gebruik word by onder meer die bestuur van 'n woordeboekprojek. Daar kan te eniger tyd bepaal word of die projek op skedule is, byvoorbeeld ingevolge tyd en die aantal toegelate bladsye. Indien die woordeboek byvoorbeeld binne twee jaar voltooi moet word en nie meer as 1 000 bladsye mag beslaan nie, beteken dit dat een blok per week afgehandel moet word en die totale lengte nie 10 bladsye per blok mag oorskry nie. Die aantal lemmas per blok en die gemiddelde lengte van die artikels kan vooraf bepaal word. Werkverrigting deur die onderskeie leksikograwe en selfs die vergoeding aan deelydse samestellers kan op dié manier gemeet word. Dit is wat Prinsloo en De Schryver (2003) met die term 'effektiewe vordering' bedoel.

Dele XII en XIII van die WAT gemeet aan die Afrikaanse Liniaal

Vir hierdie doel word die oorspronklike WAT-liniaal gebruik wat in Prinsloo en De Schryver (2003) vir die evaluering van die WAT ontwerp is, ten einde te bepaal tot watter mate die alfabetiese kategorieë P, Q en R met die voorspelde

liniaalwaardes korreleer. Figuur 8 is die bygewerkte grafiek wat ook die P-, Q- en R-waardes reflekteer.



Figuur 8: Die bygewerkte WAT-liniaal

Vir P is die liniaalwaarde 534 bladsye en die werklike aantal bladsye 718, wat ooreenkomstig die liniaal 'n matige oorbewerking suggereer. Vir Q is die liniaalwaarde 2.7 en die werklike bladsye 3 en vir R 486 en 507 respektiewelik. Kategorieë Q en R korreleer dus presies met die liniaal. Voorspelde liniaalwaardes ooreenkomstig bladsye vir die onvoltooide dele S tot W volgens die oorspronklike berekening (Prinsloo en De Schryver 2003) is S = 1 529.2, T = 598.8, U = 248.7, V = 1 061.3 en W = 523.7.

Slotopmerkings

'n Voortreflike leksikograaf moet deurgaans stry teen alle vorme van inkonsekwentheid en wanbalans tydens die samestelling van 'n woordeboek. Die oor- of onderbewerking van alfabetiese kategorieë in woordeboeke is nie bloot 'n 'tegniese' of 'akademiese' aangeleentheid nie — dit het direkte implikasies vir die woordeboekgebruiker wanneer die inligtingsaanbod verskil byvoorbeeld van te veel tot te min in dieselfde woordeboek. Leksikografiese liniale maak 'n bydrae tot die kwaliteit van woordeboeke en die gebruikersperspektief. Dit wil ook voorkom of hierdie balans maklik bepaalbaar is deurdat selfs baie klein korpusse van so min as 'n paar honderd tekens, soos in die geval van e-pos- en geselstaalkorpusse, reeds voldoende is om dié balans aan te dui.

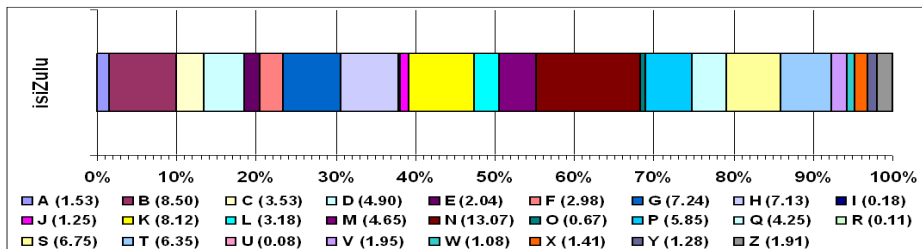
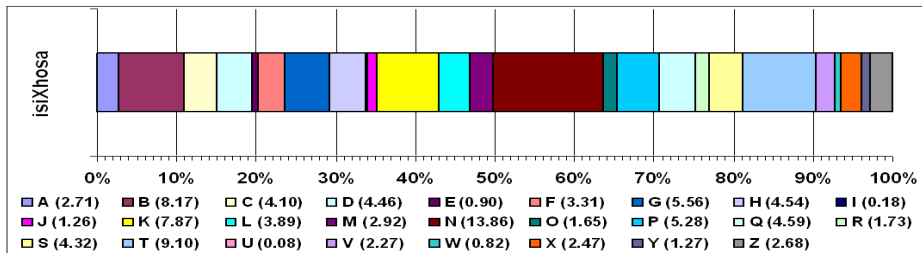
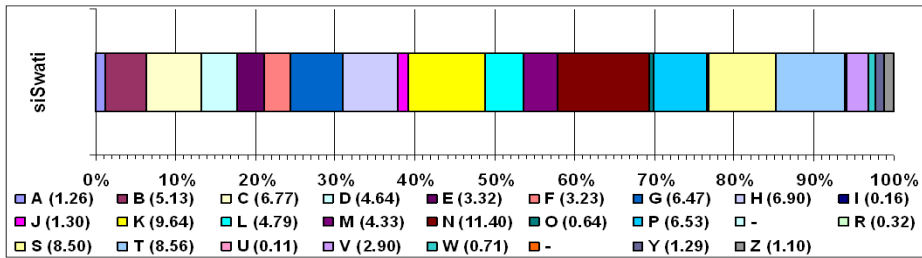
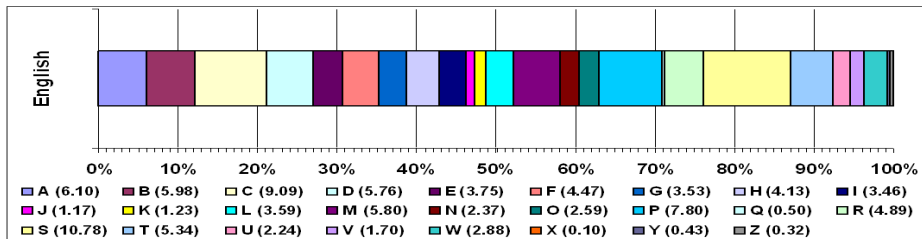
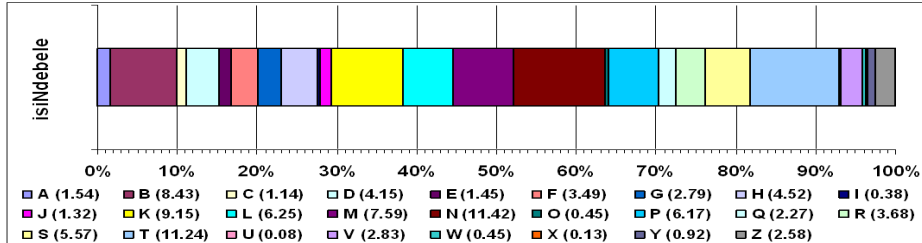
Literatuurlys

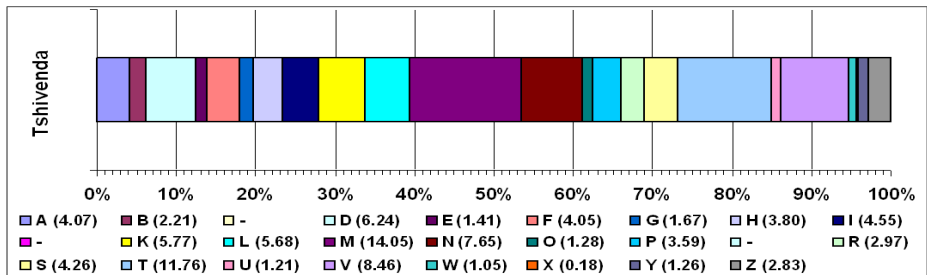
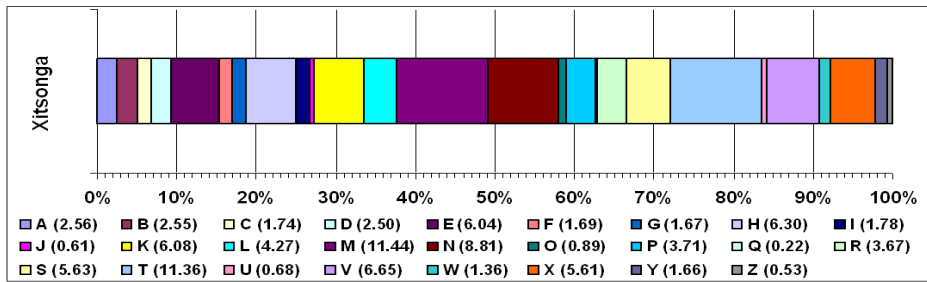
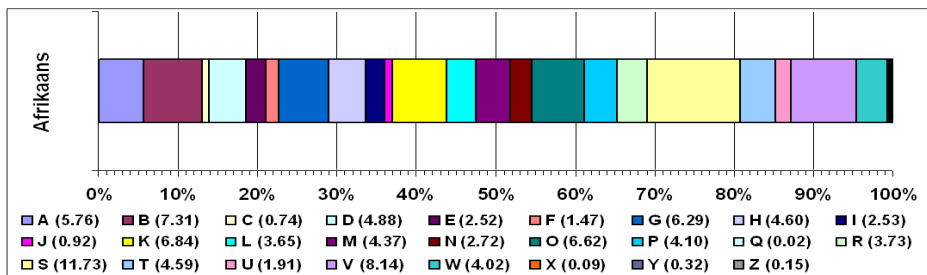
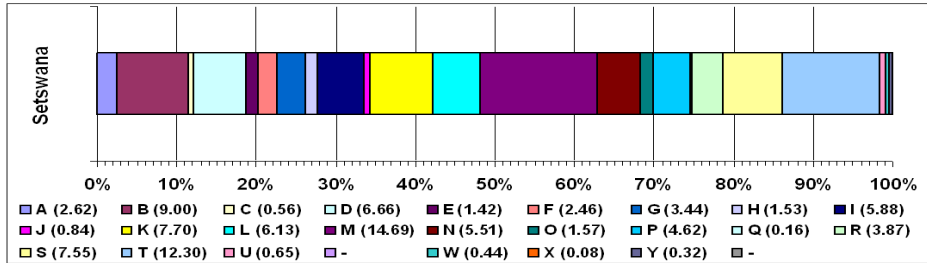
- Afrikaanse Bybel*. <http://www.bybel.co.za>
- Atkins, B.T. Sue en M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Biber, D. 1993. Using Register-diversified Corpora for General Language Studies. *Computational Linguistics* 19(2): 219-241.

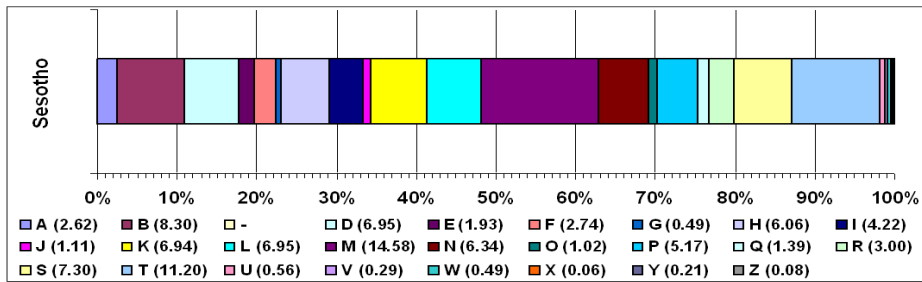
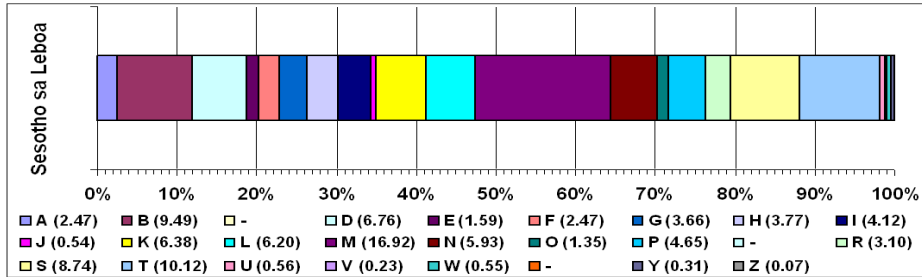
- Botha, W.** 2005. Concurrent Over- and Under-treatment in Dictionaries. A Response. *International Journal of Lexicography* 18(1): 77-87.
Brown Corpus of Standard American English. http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html.
- De Schryver, G.-M.** 2005. Concurrent Over- and Under-treatment in Dictionaries — The *Woordeboek van die Afrikaanse Taal* as a Case in Point. *International Journal of Lexicography* 18(1): 47-75.
- Eaton, J.** 1989. *Borduursteke vir Suid-Afrika. 'n Volledige gids*. Kaapstad: Delos.
- Gilbert, Z.** 1985. *Die Suid-Afrikaanse tuin. Maand vir maand*. Tweede uitgawe. Johannesburg: Central News Agency.
ICE. <http://www.ucl.ac.uk>
- Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. Londen/Nieu-York: Longman.
- Kilgarriff, A.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.
- Kriel, T.J.** 1983. *Pukuntšu Noord-Sotho-Afrikaans, Afrikaans-Noord-Sotho Woordeboek*. Pretoria: J.L. van Schaik.
- Kruyt, J.G. en M.W.F. Dutilh.** 1997. A 38 Million Words Dutch Text Corpus and its Users. *Lexikos* 7: 229-244.
Kyknet: Robinson Regstreks. DSTV. <http://www.dstv.com>
Lancaster-Oslo/Bergen Corpus. http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/LOB/lob.html.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Tweede uitgawe. New York/Cambridge: Cambridge University Press.
Landbouweekblad: Media 24.
- Langenhoven, C.J.** 1933. *Versamelde werke. Deel V*. Kaapstad: Nasionale Pers.
- Langenhoven, C.J.** 1934. *Versamelde werke. Deel VI*. Kaapstad: Nasionale Pers.
Loslyf. Afrikaanse sekstydskrif. Oktober 2000. 6(5).
Media24. <http://www.media24.co.za>.
- Otlogetswe, T.J.** 2007. *Corpus Design for Setswana Lexicography*. Ongepubliseerde Ph.D.-proefskrif. Pretoria: Universiteit van Pretoria.
Pharos-toetskorpus. <http://www.pharos.co.za>
- Prinsloo, D.J.** 2006. Compiling a Bidirectional Dictionary Bridging English and the Sotho Languages: A Viability Study. *Lexikos* 16: 193-204.
- Prinsloo, D.J. en G.-M. de Schryver.** 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with Special Reference to Afrikaans and English. Braasch, A. en C. Povlsen (Reds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 13-17, 2002*: 483-494. Kopenhagen: Sentrum vir Taaltegnologie, Universiteit van Kopenhagen.
- Prinsloo, D.J. and G.-M. de Schryver.** 2003. Effektiewe vordering met die *Woordeboek van die Afrikaanse Taal* soos gemeet in terme van 'n multidimensionele Liniaal. Botha, W. (Red.). 2003. *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*: 106-126. Stellenbosch: Buro van die WAT.
- Prinsloo D.J. en G.-M. de Schryver.** 2005. Managing Eleven Parallel Corpora and the Extraction of Data in all Official South African Languages. Daelemans, W., T. du Plessis, C. Snyman en L. Teck (Reds.). 2005. *Multilingualism and Electronic Language Management. Proceedings of the 4th*

- International MIDP Colloquium, 22–23 September 2003, Bloemfontein, South Africa*: 100-122.
Studies in Language Policy in South Africa 4. Pretoria: Van Schaik.
- Rapport: Media 24*.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Summers, Della.** 1993. Longman/Lancaster English Language Corpus — Criteria and Design. *International Journal of Lexicography* 6(3): 181-208.
- Tshwanelex*. <http://tshwanedje.com/tshwanelex/>
- Turpin-Delport, L.** 1988. *Die Suid-Afrikaanse boek van laslappie- en appliekwerk*. Kaapstad: C. Struik.
- UP Strategies Plan*. 2002–2005. <http://www.up.ac.za>
- Woordeboek van die Afrikaanse Taal, Deel XII (P–Q)*. 2005. Botha, W.F. (Hoofred.). Stellenbosch: Buro van die WAT.
- Woordeboek van die Afrikaanse Taal, Deel XIII (R)*. 2009. Botha, W.F. (Hoofred.). Stellenbosch: Buro van die WAT.

Bylaag A: Liniale vir die 11 amptelike landstale (Prinsloo en De Schryver 2005)







Checking Knowledge in Online Encyclopaedias: Towards a Behavioural Approach to Data Accessibility

Marcin Overgaard Ptaszynski, *Centre for Lexicography, Aarhus School of Business, Aarhus University, Denmark (maop@asb.dk)*

Abstract: This article outlines a new approach to data access in electronic encyclopaedias. Contrary to most of the existing research, which treats access as a lexicographer-dominated, structural concept, access is viewed here as a process which is essentially user-driven. The role of the lexicographer is to facilitate this process, so that it can become rapid and unimpeded, but their control of it should be reduced to a minimum. This can only be achieved if access to data is adjusted to the specific type of need for information. Accordingly, the article examines three forms of data access applicable in situations in which the user experiences the need for a finite, specific amount of information. The results of the examination confirm that access to data is indeed a process whose success depends on the interaction between the user, the data, and the type of need for information which the given reference work aims to satisfy.

Keywords: ENCYCLOPAEDIA, ELECTRONIC LEXICOGRAPHY, ACCESS, STRUCTURE, ACCESS ROUTE, SIMPLE INFORMATION NEED, COMPLEX INFORMATION NEED, KNOWLEDGE ACQUISITION

Opsomming: Die nagaan van kennis in aanlynensiklopedieë: Op weg na 'n gedragsbenadering tot datatoeganklikheid. Hierdie artikel skets 'n nuwe benadering tot datatoegang in elektroniese ensiklopedieë. In teenstelling met die meeste bestaande navorsing wat toegang behandel as 'n leksikograaf beheerste, strukturele konsep, word toegang hier gesien as 'n proses wat hoofsaaklik gebruikersgedrewe is. Die rol van die leksikograaf is om hierdie proses te vergemaklik sodat dit vinnig en ongehinderd kan plaasvind, maar sy beheer daarvan behoort tot 'n minimum beperk te word. Dit kan slegs bereik word indien toegang tot data aangepas word by die spesifieke soort behoefte aan inligting. Gevolglik behandel die artikel drie maniere van datatoegang wat van toepassing is in situasies waarin die gebruiker die behoefte vir 'n definitiewe, spesifieke hoeveelheid inligting ondervind. Die resultate van die ondersoek bevestig dat toegang tot data inderdaad 'n proses is waarvan die sukses berus op die wisselwerking tussen die gebruiker, die data en die soort behoefte aan inligting wat die bepaalde naslaanwerk beoog om te bevredig.

Sleutelwoorde: ENSIKLOPEDIE, ELEKTRONIESE LEKSIKOGRAFIE, TOEGANG, STRUKTUUR, TOEGANGSROETE, EENVOUDIGE INLIGTINGSBEHOEFTE, KOMPLEKSE INLIGTINGSBEHOEFTE, KENNISVERKRYGING

1. Introduction

What is the wingspan of an albatross? Anyone who would like to know the right answer may ask an expert in the field, e.g. an ornithologist or a biologist. If none of them happens to be available and the issue needs to be resolved without delay, one may consider consulting a reference work. The decision to do so reflects the fact that one has a lexicographically relevant need. The need to obtain information about the wingspan of an albatross can be described more generally as a need to acquire knowledge about a specific subject matter.¹ It can be satisfied by means of an appropriate reference work containing relevant data, in this case an encyclopaedia, a bird atlas, or perhaps a dictionary of ornithology.

Offering the user the correct and relevant data is, however, only one of the objectives of lexicography. A great number of reference works do contain the data that satisfy the user's particular need for information, but searching and finding these data may be difficult and time-consuming. This may discourage the user from further consultation of the particular reference work. Indeed, the satisfaction of a need for information depends not only on the presence of specific data, but also on their being accessible in a way that requires the minimum time and effort on the part of the user.

This article investigates access to data in electronic encyclopaedias. The focus is on the satisfaction of the user's need to obtain a precise answer to a specific question rather than to acquire extensive knowledge about a given subject. A brief presentation of access routes available in electronic encyclopaedias and an examination of the notion of access in lexicographical literature reveal that access is customarily conceived of as structures, which may be optimised to allow more effective satisfaction of information needs. This view is challenged here by considering access in terms of a process. The validity of such an approach is subsequently verified by testing three different forms of access to data.

2. Simple and complex needs for information

The need to acquire or revise knowledge about a specific subject matter has been recognised as lexicographically relevant by scholars representing various theoretical approaches to lexicography. For instance, Tarp (2008) calls it a "cognitive user need" as distinct from a "communicative user need", which is understood as a need for linguistic assistance in an act of communication. This division mirrors the one between addressing general issues ("globale Fragestellungen") and specific issues ("punktuelle Fragestellungen"), proposed much earlier by Hausmann (1977: 144). Both kinds of issues are related here to the area of language, but there is a significant difference between them. Addressing general issues reflects the need to gather factual knowledge, such as "How

is the vocabulary of a particular language structured?" or "What is the lexical stock of language X that deals with the specific subject Y?". In contrast, addressing specific issues reflects the need for a solution to a problem arising in an act of communication, and can be exemplified by such questions as "What is the meaning of X?" or "What is the equivalent of X in language Y?"

From the lexicographical point of view, it does not seem unreasonable to distinguish the need to acquire or revise knowledge from the need to solve communication-related problems. In terms of reference works used to satisfy both kinds of need, the former usually require an encyclopaedia or an encyclopaedic dictionary, while the latter are customarily addressed by various types of general-purpose and specialist dictionaries (although, certainly, both kinds of dictionaries can be used to satisfy cognitive user needs too by offering data that convey information about language or specialised information about a specific topic). More important, however, are the nature and complexity of issues that characterise either kind of need. Looking for a solution to a problem connected with text production, text reception or translation is necessarily confined to linguistic activity and clearly exemplifies a need for specific information. By contrast, gathering knowledge about a given subject is by no means restricted to the area of linguistics and may happen in response to either a general or a specific problem. This assumption and the implication it has for access to data constitute the point of departure for this discussion.

Whatever the area of knowledge one is exploring, one's need for information will fall into one of two broad categories, determined by the degree of complexity of the issue addressed. A simple need is one that can be satisfied by means of a precise, specific, finite amount of information. For example, if one needs to know when the Federal Republic of Germany was proclaimed, a single piece of data will be sufficient: 23 May 1949. In contradistinction to this, if one would like to know how the Reichstag functions, the need for information is complex. It cannot be satisfied by means of a specific piece of data placed in a particular article in a given encyclopaedia. Solving problems of this kind often requires a systematic consultation of one or more reference works.

The distinction between simple and complex needs has a direct bearing on access to data that satisfy them. This is particularly relevant with regard to the needs for information about well-known people, geographical locations, or central concepts within specific domains, such as *function* within mathematics. In many encyclopaedias, articles dealing with such issues contain considerable amounts of data. If one's need for information is simple, e.g. one is only interested to know of what Ferdinand Porsche died, then the prospect of having to peruse all other information about Porsche before one learns what caused his demise might easily put one off using the particular reference work. The relevant data must be reached more efficiently. This confirms that there is a crucial difference between simple and complex needs, but, as will be revealed later in this article, it has not been taken into consideration by metalexigraphers.

3. Access routes in electronic encyclopaedias

Data stored in electronic encyclopaedias (and other reference works designed to help the user with the acquisition of knowledge) can be accessed in a variety of ways. Four basic types of access routes can be distinguished, depending on their starting point, i.e. the place in or outside the given reference works, where the user begins the search for information.

First, data can be searched by means of the list of headwords. All items on the list contain hyperlinks, which, when activated, lead the user directly to the respective articles. This type of access route is available in most encyclopaedias on CD-ROM or DVD-ROM, such as *Britannica 2006 CD-ROM* (2006), as well as in some internet-based encyclopaedias, e.g. *Wikipedia* (2009)² and *Aschehougs Leksikon* (2009).

Second, the user can access data via an index of topics or a portal. All items in the index and some items in a portal contain hyperlinks, which, when activated, lead the user to particular articles, either directly or through other, more specific indexes or portals. This type of access route is found in encyclopaedias with a partly thematic access structure, e.g. *Wikipedia* (2009), *Quid* (2009), and *Den Store Danske* (2009). A special type of index of topics is the so-called 'entry menu'. It is an interactive table of contents placed within an article. Obviously, accessing data via entry menus presupposes opening the relevant article in the first place.

Third, data can be accessed through other hyperlinks. These may be placed inside articles, in the outside matter, or outside the given reference work, for instance on an unrelated website or in other software. They can direct the user to particular articles, to the outside matter, or to data placed outside the encyclopaedia. Hyperlinks can be overt, i.e. the format of a particular piece of data (and often the shape of the cursor if placed in the vicinity of the data) immediately make the user aware of the presence of a hyperlink in it. By contrast, covert hyperlinks are invisible to the user. Some encyclopaedias, like *Wikipedia* (2009), offer only overt hyperlinks, others, like *Treccani* (2009), contain only covert hyperlinks,³ while still others, such as *Larousse* (2009), offer a combination of both kinds.

Fourth, nearly all reference works allow their users to access data through a search box. It is usually possible to narrow the scope of the search by modifying it in a variety of ways. For instance, the search can be limited to a specific part of the encyclopaedia, such as the list of headwords, as in *Wikipedia* (2009), a media collection, as in *Larousse* (2009), an atlas, as in *Uniwersalna Encyklopedia PWN* (2006), or a specific topic, as in *Chronik der Weltgeschichte* (2002). Also, most modern encyclopaedias allow some kinds of Boolean search in order to limit the quantity of data presented as the search result. In very few reference works, such as *Musikorbogen* (2009), the user is additionally obliged to specify the need that has triggered the consultation, e.g. problems with text reception, or the desire to acquire more knowledge about a given topic. Submitting a

query in a search box may yield various results. If the search string matches a headword in the reference work, the respective article is usually opened. Otherwise, a list of relevant headwords is displayed, sometimes with extracts from the respective articles. The extracts normally do not vary according to the search string submitted (exceptions include *Wikipedia* (2009), and other wikis, as well as *Quid* (2009) where the search is powered by Google). The results of searching within other types of data than text may include a list of photographs, films or maps, as in *Larousse* (2009).

The four types of access routes have been analysed and described, both in lexicographical literature and in numerous reference works. Still, very few of them seem to have been devised with a deliberate view to satisfying either simple or complex needs for information. In most cases, the route the user follows terminates in a particular article presented in its entirety, regardless of whether this is necessary or not to satisfy the given need for information efficiently.

The structures provided by the lexicographer neither constitute nor guarantee in themselves rapid and unimpeded access to data if they are not devised with the specific needs and competences of the prospective users in mind. Accordingly, the importance of the relationship between structures and user needs and skills is increasingly recognised by lexicographers (see, e.g. Bergenholtz and Gouws 2007, Gouws 2009). The relevant studies, however, mostly concern dictionaries, often in printed form. Research into needs-adapted access to data in electronic encyclopaedias, especially such that would shed more light on the process taking place during the interaction between the user and the particular structure, has yet to appear.

4. Research on access to data in electronic reference works

Access to data in electronic reference works has enjoyed wide coverage in literature on lexicography. The research on the subject is characterised by four major features. It is mostly related to dictionaries rather than to encyclopaedias. It is usually done as part of broader studies in electronic lexicography. It seldom explores the relation between access to data and specific types of needs for information. Finally, it often represents an approach which is inadequate for its purpose. Each of these features is examined in more detail below.

Discussion of access to data in electronic reference works usually focuses on dictionaries whose aim is to satisfy the user's need for immediate linguistic assistance. Information concerning data access in electronic encyclopaedias and dictionaries aiming to help the user with knowledge acquisition is much rarer. It usually appears in project reports (Pedersen and Rasmussen 2007, Biffi 2009), reviews (Beißwenger 2002, Eickmeyer 2002, Lobenstein-Reichmann 2002), and comparative analyses of selected reference works (Streitberger 2002).

From the above it can be surmised correctly that access to data is often approached from a practical, descriptive standpoint and usually in connection

with other aspects of electronic lexicography. Theoretical studies whose primary subject is access to data in electronic reference works are rare (though see Bergenholtz and Gouws 2007, Tono 2009, Tarp 2009). To come across theoretical discussions concerning access, one often needs to consult research of a broader scope, such as general descriptions of electronic dictionaries (Nesi 1999, De Schryver 2003) or studies in digitisation of older reference works (Lobenstein-Reichmann 2007).

The scarcity of theoretical studies may explain why few metalexicographers hitherto have shown any interest in the relation between access to data on the one hand and particular types of need for information on the other. Establishing solid theoretical links between these two phenomena belongs to exceptions (Bergenholtz and Gouws 2007, Tarp 2009). Occasionally, one can witness a proposal for some basic principles for a normative theory, as in Nielsen and Mourier (2005). In some other studies, specific types of information needs are not mentioned explicitly but may be deduced from the context. For example, what Beißwenger (2002: 318) surely has in mind when he proposes suggestions for optimising data access for a user who "über das geistige Leben im Hochmittelalter informieren möchte" must be complex needs for information arising in connection with knowledge acquisition. Yet other researchers take the need for information for granted and do not speculate about its type, even though they recognise the importance of securing quick and easy access to data. Thus writes e.g. Streitberger (2002: 99), explaining how users can quench their "kunstgeschichtlicher Wissensdurst" with *Belser Lexikon der Kunst- und Stilgeschichte* (1999). Finally, some metalexicographers reject the issue of adapting data access to particular information needs, as they believe the problem to be insoluble. Trap-Jensen (2005: 115), for example, claims that owing to practical constraints it is simply not possible to devise search options in reference works on the basis of specific types of need for information.

Perhaps the most interesting feature of research on access to data in electronic reference works is the fact that it makes use of theoretical notions which are not quite adequate for its purpose. The definition of the very term *access* is a case in point. As the following examples clearly illustrate, in lexicography, access is commonly perceived in terms of search routes and structures in a reference work, which users follow and enter in order to find data:

The access structure can be regarded as the search route the dictionary user follows during a dictionary consultation procedure. (Steyn 2004: 276)

Die Datenakzessivität betrifft die Zugriffsbereitschaft und damit die Nachschlagbarkeit von textuellen lexikographischen Daten. Grundsätzlich gilt: Es muss zwischen der externen und der internen Datenakzessivität unterschieden werden. Weiterhin gilt: Es muss die nichtmediostrukturelle von der mediostrukturellen Datenakzessivität unterschieden werden. (Wiegand 2008: 214)

Such a conception of access represents patterns of cognitive ergonomics seen from the lexicographer's perspective. For the user, it is not only important to

know what search route to follow, but also how to formulate their query, what search result to expect, and, possibly, how much time the entire procedure will take. Moreover, the validity of this approach is mostly confined to printed reference works. For users of electronic lexicographical resources, especially internet-based ones, the mediostructure is not necessarily significant. Data displayed on screen are often retrieved from relational databases, rather than reached via predefined routes through macro-, medio- and microstructures. Certainly, electronic reference works do have access structures, too. However, the preconceptualised search route which underlies the access structure need not be followed by the user during a consultation procedure. The actual search route is frequently through databases and relations between them and is constructed by the user solely for the purpose of a specific act of consultation. Indeed, this is the essence of the electronic revolution, which "lies in the fact that users are liberated from the alphabetical straitjacket, that hypertext, menus, etc. eliminate (artificial) linear text restrictions, that the data conjured up onscreen are not static, and that powerful search capabilities ensure a smooth overarching navigation" (De Schryver 2003: 157). The systems of particular "Leitelemente" (Wiegand 1988) and their mutual relations, devised by the lexicographer, seem to have a diminished significance for the user of an electronic reference work, although it still can be argued that the lexicographer simply does not have to design such relations — they can exploit the possibilities offered by the electronic media.

From the above it appears that *access* is a term denoting a kind of process, where the structure is just one of the elements involved. This view is shared by Bergenholtz and Gouws (2007: 240), who speak of "the notion of an access process [...] which includes the access structure as one of its components". This dynamic view is quite convincing, as it can actually be extended so that access does not just denote an instrumental process, but also a constructionist one, defined as follows:

Access is a process in which the user of a reference work deliberately seeks and finds data which immediately satisfy their specific, lexicographically relevant needs. The process is user-initiated, user-motivated, and user-driven, and it utilizes inner and outer structures of the reference work. It begins with the user's selection of the reference work and their formulation of the query, it makes use of a specific search route, and it terminates once the user has reached the required data, which are found among the search results provided in response to the query.

The process invariably makes use of specific external and/or internal search routes, which are partly or entirely devised by the lexicographer but activated solely by the user. Access is thus observable and thereby amenable to empirical investigation. This can come in the form of both direct observation, including eyetracking analysis, and remote, asynchronous investigation, such as an analysis of log files registering the use of an electronic reference work. The process is also measurable in time, since one can specify its beginning and its

end. The term *access* thereby covers such terms as *search route*, *search string*, *search result* and *search time*.

The validity of the definition proposed above will be tested in the next section. It contains an examination of various forms of accessing data in response to a simple need for information.

5. Access as a constructionist interaction

To argue in favour of the conception of access as a user-driven process, a short case study is made and presented below. Its aim is to illustrate what happens when the user accesses data in order to satisfy a simple need for information in connection with knowledge acquisition.

5.1 Methodological considerations

The case study consists in testing access to data via three types of access routes which appear to be most appropriate for the retrieval of specific information. These include reaching data through the entry menu placed in the article, through a Boolean search performed in a search box, and through a full text search of HTML text performed in the user's browser. Other types of access routes are disregarded here, each for a different reason. Accessing data solely via a list of headwords does not seem to be particularly suitable for satisfying simple needs for information (cf. the problem of Ferdinand Porsche's death mentioned earlier). Following a hyperlink to a specific article from outside the reference work, e.g. through performing a Google search, does not reflect the conscious choice of that work to satisfy one's need for information. Finally, attempting to reach data through a hyperlink from one article to another is more likely to reflect a stage in satisfying one's complex need for information.

All the tested attempts at accessing data are done in response to a realistic need for specific information.⁴ The need can be formulated as the following question: Of what did Frédéric Chopin die? It should be kept in mind, however, that the objective of the test is not to reveal the presence or absence of particular data in a given reference work but to prove certain assumptions about access to data. The data themselves are of secondary importance, as is the truth value of the information they convey.

Only one reference work is subjected to investigation. The choice of *Wikipedia* (2009) for this purpose seems reasonable for several reasons. It is currently the world's biggest general-purpose encyclopaedia for laypeople, it is available in a number of languages, and, being an internet-based reference work, it allows the user to access data via all the routes to be investigated. At the same time it must be borne in mind that *Wikipedia* (2009) is only used here to illustrate the points made. It is not meant as an example to be followed by other electronic encyclopaedias, nor as the most popular or effective reference

work whose use reflects user-initiated search strategies. Users might try to satisfy their information needs more frequently by Google. Answer to quiz-like questions may be sought in specialist reference works, like *Answers.com* (2010). Therefore, wherever applicable, references will be made to other electronic encyclopaedias.

The case study is carried out in controlled conditions. It is a qualitative investigation. There is only one informant involved, and it is the author (hence the frequent use of the first person pronoun in the following sections). The informant is an interested layperson in the area of knowledge to which the information need specified above pertains. Although studies in access to data, which involve only one informant are not uncommon (see e.g. Bergenholtz 2009), there are obvious limitations to what they can reveal. To be of scientific significance, their results must be verified in an experiment involving more informants. The study is thus best regarded as part of a hypothesis formulation, which needs to be tested by a proper experimental procedure.

5.2 Entry menu

Before being able to reach data through an entry menu placed in an article in *Wikipedia* (2009), two other actions need to be performed. First of all, having opened the start page, one must choose the language of the encyclopaedia, and then open the relevant article. I choose the English version. To find out about the cause of Chopin's death, I type "Chopin" in the search box placed in the left menu. Typing a query in the box and clicking the button "Go" will activate a search in the list of articles. Once I type, the type-ahead search is being automatically activated, as a drop-down menu pops up beneath the box, suggesting articles whose headwords begin with the search string. I decide to choose "Chopin". This opens the article, which happens to be of considerable length (slightly over 11 000 words). Only the first section of the article is displayed on screen, the remaining text becoming visible as I scroll down. Not knowing where the data satisfying my need for information are located in the article, to save time I turn to the entry menu. It is rather concise, consisting of twenty items: nine main headings (*Life, Memorials, Music, Works, Fiction, See also, Notes, Bibliography, External links*) and two sets of altogether eleven subheadings (six under the heading *Life* and five under the heading *Music*) arranged at the second level of nesting (i.e. 1.1.). I click on the subheading 1.6. *Final years*, upon which I am redirected to the respective section in the article. The section features two graphic images and a longish text of nearly 900 words. I need to read the entire text, or at least scan it carefully, down to the very last two lines, from which I finally learn that "[i]n 2008, a controversy arose over whether Chopin died of tuberculosis or cystic fibrosis [...]." The procedure is rather time-consuming, which makes me wonder whether the same data could be reached more efficiently.

Could the problem possibly have been avoided if the entry menu had been more fine-grained and each item had led to shorter chunks of the article? This cannot be determined with any certainty. The form and content of the entry menu were adequate for my user profile. My prior knowledge about Chopin was restricted to some very basic facts concerning his birth, his death, his nationality, and his love affair with George Sand; my knowledge of his music was and is far from expert. This was nearly mirrored by the simple and concise menu. The fact that it led me to a lengthy text may have dashed my hopes of reaching the data quickly and conveniently, but it did not reveal any problems with the structure of the menu. This was confirmed when I tried to use the French *Wikipedia* (2009) article about Chopin to satisfy the same need for information. The portions of text displayed on activating an item from the entry menu are usually shorter than they are in the English article. However, the menu itself consists of no less than 101 items, spread over five levels of nesting (1.1.1.1.1.). Its usefulness presupposes the user's expert knowledge about Chopin, which I do not happen to have.

The fact that retrieving specific information via an entry menu may not live up to the expectations of some users is perhaps interesting to critics of encyclopaedias. More important, however, is what it reveals about the process of accessing data. The procedure is likely to fail because an interaction between the user and the structure it involves is to a high degree motivated by the lexicographer instead of the user. To begin with, the user has to guess which article to open in order to satisfy their need for information. Whether their guess is correct or not depends in principle on the lexicographer's whims about where to place the relevant data. Next, they need to know which item from the entry menu is most likely to lead them to the required data. This can be much more of a challenge than choosing the correct article. An entry menu invariably represents a prefabricated structuring of information on the basis of some thematic or ontological arrangement of ideas which the lexicographer has related to the notion expressed by the headword. To be of any practical use, the form and contents of the menu must largely correspond to the amount and arrangement of information about the given concept in the user's mind. For obvious reasons, this requirement is unlikely to be met with regard to all intended users of a general-purpose encyclopaedia. Finally, on activating the correct item from the menu, the user would reasonably expect to be redirected to a text in which they can locate the relevant data quickly and conveniently. The length and wording of the text, however, reflect the choices made by the lexicographer, not by the user.

Clearly, placing an entry menu in an article, the lexicographer undoubtedly wishes to help the user retrieve specific information. If rapid and unimpeded data access was just a matter of equipping the reference work with appropriate structures, this solution should always work. That this is not the case confirms that access to data is a process and that any constraints placed on the user's control of it are likely to compromise its success.

5.3 Boolean search

The start page of *Wikipedia* (2009) does not feature a search box which supports Boolean searching in the text of all articles. Such a search box does, however, appear in the left menu when the user has chosen the language of the encyclopaedia. I select English, and, determined to find out quickly and easily what caused Chopin's death, I type "Chopin + died" in the box and click the button "Search". In response to this, the search engine finds all the 783 articles containing both keywords and sorts them by relevance. Extracts from the first twenty articles are displayed on screen. Each one is introduced by its respective headword and contains the key words in context, set in boldface for emphasis. The first article listed among the search results is "Frédéric Chopin". The extract displayed beneath the headword contains occurrences of both "Chopin" and "died", but it can only inform me that Chopin died aged thirty-nine, and that his friend Cherubini had died aged eighty-one in Paris in 1842. This does not satisfy my need for information.

From this point, I can follow two possible routes to reach the relevant data. I can open the article about Frédéric Chopin and look for the data there anyway. Finding them might take time, regardless of whether I choose to read the whole article or use the entry menu instead. It would be of considerable help if all occurrences of the keywords were emphasised inside the article text,⁵ but unfortunately this is not the case. Therefore, I decide to abandon the article and examine whether the relevant data appear in the extracts belonging to some other articles. I do not find them in the extracts from the three articles immediately following the first result. However, the fifth result on the list, the article about the role-playing video game *Eternal Sonata*, does contain data satisfying my need for information. The extract informs me that "the game is centered on the Polish romantic pianist and composer Frédéric **Chopin**, who **died** of tuberculosis at the age of 39. [...]". Since I have never heard about the game before, I would not have looked for the answer to my question in the article about it.

The procedure described above provides further incitement to conceive of access to data as a process rather than a structure. It also confirms that the process is user-driven and user-motivated. To satisfy their need for specific information quickly and easily, the user does not have to know which article to open, though, naturally, a certain level of (especially linguistic) competence will facilitate a quick choice. Ideally, they do not have to open any article, as the relevant data are already displayed in one or more extracts featured among the search results.⁶ This renders both the composition of the article and the structures the lexicographer has provided to navigate quickly in it totally insignificant.

Admittedly, the user must be lucky in their forming of the search string. While searching for "Chopin + died" in *Wikipedia* (2009) would produce a positive result, submitting "Chopin + death", "Chopin + dead" or "Chopin + demise" would not. The interaction between the user and the structure is only

successful provided the keywords appear in any article in the reference work. On the other hand, even though it does have a bearing on the outcome of the process, the degree of the lexicographer's control of the interaction is in fact very low. The decision concerning the form of the search string resides ultimately with the user.

5.4 Full text search of the HTML text

Since *Wikipedia* (2009) is internet-based, and all its articles are displayed as HTML texts, there exists yet another way of quickly and easily retrieving specific information from it. On the start page of the encyclopaedia, I choose the language (English) and then open the article "Frédéric Chopin", hoping that it features the data which will satisfy my need for information, which is to learn what caused Chopin's death. Once the article is displayed on screen, I hit two keys on the keyboard: *Ctrl* and *F*. This activates the search window in my Internet browser (Internet Explorer). Into it I type "died", to which the browser replies by highlighting all occurrences of the search string in the article text.⁷ The first occurrence is found towards the end of the lengthy introductory section of the article in the following context: "Always in frail health, he **died** in Paris in 1849, aged thirty-nine, of pulmonary tuberculosis." It happens to satisfy my need for information. If this had not been the case, I would have examined the next occurrence of "died", to which I would have been redirected by clicking on the button "Next" in the browser's search window.

Certainly, the success of this form of data access depends on the correct choice of the article which features the relevant data, and this, in turn, has been determined by the lexicographer. However, once the article is displayed on screen, the user's reliance on the structures provided by the lexicographer diminishes. Actually, the structures are abandoned in favour of a solution provided by the programmer of the Internet browser. The lexicographer certainly still deserves some of the credit if they have incorporated the solution deliberately with a view to facilitating the access process for the user. On the other hand, if this is the case, one may wonder why performing text search of the HTML document is not mentioned in user's guides to online encyclopaedias,⁸ together with other instructions on how to search for information. If the user is not informed of this possibility, chances are that they successfully construct their own access process by exploiting the medium of the reference work regardless of whether that was the intention of the lexicographer. Such cases cannot be accounted for if access to data is discussed solely in terms of lexicographical structures.

6. Final remarks

The objective of this article was to provide a new insight into access to data in electronic reference works. It was argued that access should not be viewed as a

structural, instrumental notion determined by the lexicographer, but as a process initiated, constructed and controlled by the user.

The efficiency of this process depends on two factors. First, access to data needs to be adapted to the specific type of the user's need for information. The examination of three forms of access to data carried out in the article was restricted to the issue of satisfying simple needs for information. If the user's need had been a complex one, the process of accessing data would have been different. This matter requires advanced, experimental research into how users precisely behave when they look for data in specific settings. Search behaviour must be regarded as a cognitive interaction, and future research must include psychological manifestation of this cognitive activity. This has been revealed, although to a very modest extent, in the case study. Second, reference works must be constructed in a way that allows the user maximum control of the process of accessing data, including adaptive design.

The reference works referred to in the article were general-purpose encyclopaedias for laypeople. It should be kept in mind that access to data in connection with the acquisition of knowledge is an area of research that should not be limited to the genre investigated. Satisfaction of the same lexicographically relevant needs may be sought in other reference works, such as specialised dictionaries and handbooks, which account for an overwhelming majority of reference works produced, but still attract relatively little attention of metalexicographers. Investigating such works may reveal more insights into the relation between access to data, information needs of the user, and particular lexicographical genres aiming to satisfy those needs.

Endnotes

1. Although a layperson might experience the need for this information, the exact question asked is likely to differ from the one presented here. As *wingspan* might not be part of the vocabulary of such user, a more probable form of the question is "How big is an albatross?" Here, however, for the sake of precision the question is formulated in such a way as to clearly indicate the precise need for information.
2. Admittedly, this access route is not the one most commonly used in *Wikipedia* (2009).
3. In *Treccani* (2009) the user is actually made aware of the fact that all words in the article contain hyperlinks. A bar featuring this information always appears above the displayed article.
4. This excludes information needs based on false presuppositions, such as the need to know the name of Nicolas Sarkozy's sister (he has no sister), and derived from patently self-contradictory premises, such as the need to know the place of death of the Sultan of the Vatican Republic who ruled in the period 3045–3010 AD.
5. This feature is found in e.g. *Aschehougs Leksikon* (2009) and *Den Store Danske* (2009).
6. This requires, however, that the extracts always vary according to the search string submitted.
7. This applies also to Google Chrome and Mozilla Firefox.

8. The only counterexample known to the author is that of the former online version of the Swedish *Nationalencyklopedin*. The user's guide to the current version does not contain the information on accessing data via full text search of an article.

References

Reference works

- Answers.com. The world's leading Q&A site.* <http://www.answers.com> [accessed 4 August 2010].
- Aschehougs Leksikon.* <http://www.aschehougsleksikon.dk> [accessed 4 December 2009].
- Belser Lexikon der Kunst- und Stilgeschichte.* Version 2.0. Munich: USM.de Systemor, 1999.
- Bergenholtz, I.** 2009. *Musikorbogen.* Århus: Handelshøjskolen i Århus. <http://www.musikorbogen.dk> [accessed 4 December 2009].
- Britannica 2006 CD-ROM = Encyclopaedia Britannica 2006 Deluxe Edition CD-ROM.* London: Encyclopaedia Britannica Inc., 2006.
- Chronik der Weltgeschichte Version 3.0. Der neue Kulturfahrplan nach Werner Stein.* CD-ROM. Munich: USM.de Systema, 2002.
- Den Store Danske. Gyldendals åbne encyklopædi.* <http://www.denstoredanske.dk> [accessed 4 December 2009].
- Larousse = Encyclopédie contributive Larousse.fr.* <http://www.larousse.fr/encyclopedia> [accessed 4 December 2009].
- Nationalencyklopedin.* <http://www.ne.se> [accessed 5 March 2010].
- Quid = L'Encyclopédie Quid.* <http://www.quid.fr> [accessed 4 December 2009].
- Treccani — Enciclopedia e Vocabolario on line, Articoli, Video.* <http://www.treccani.it> [accessed 4 December 2009].
- Uniwersalna Encyklopedia PWN = Uniwersalna encyklopedia multimedialna PWN.* (Wersja 1.) Warsaw: PWN, 2006.
- Wikipedia. The Free Encyclopedia.* <http://www.wikipedia.org> [accessed 4 December 2009].

Other literature

- Beißwenger, M.** 2002. (Review of) *Chronik der Weltgeschichte Version 3.0.* *Lexicographica* 18: 312-320.
- Bergenholtz, H.** 2009. Hurtig og sikker tilgang til informationer om ordforbindelser. *LexicoNordica* 16: 29-54.
- Bergenholtz, H. and R.H. Gouws.** 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica* 23: 237-260.
- Biffi, M.** 2009. Accademia della Crusca's Online Dictionaries. Bruti, S., R. Cella and M.F. Albert (Eds.). 2009. *Perspectives on Lexicography in Italy and Europe:* 239-286. Newcastle upon Tyne: Cambridge Scholars Press.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Eickmeyer, J.** 2002. (Review of) *Reclams elektronisches Romanlexikon. Deutschsprachige erzählende Literatur vom Mittelalter bis zum Gegenwart.* *Lexicographica* 18: 311-312.

- Gouws, R.H.** 2009. Sinuous Lemma Files in Printed Dictionaries. Access and Lexicographic Functions. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 3-21. Amsterdam/Philadelphia: John Benjamins.
- Hausmann, F.J.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Romanistische Arbeitshefte 19. Tübingen: Max Niemeyer.
- Lobenstein-Reichmann, A.** 2002. (Review of) Lexikon des Mittelalters. *Lexicographica* 18: 300-301.
- Lobenstein-Reichmann, A.** 2007. Allgemeine Überlegungen zur Retrodigitalisierung historischer Wörterbücher des Deutschen. *Lexicographica* 23: 173-198.
- Nesi, H.** 1999. A User's Guide to Electronic Dictionaries for Language Learners. *International Journal of Lexicography* 12(1): 55-66.
- Nielsen, S. and L. Mourier.** 2005. Internet Accounting Dictionaries: Present Solutions and Future Opportunities. *Hermes* 34: 83-116.
- Pedersen, J.H. and T.E. Rasmussen.** 2007. Strategi i navigation. *LexicoNordica* 14: 71-80.
- Steyn, M.** 2004. The Access Structure in Learner's Dictionaries. *Lexikos* 14: 275-298.
- Streitberger, A.** 2002. Wörterbücher der Kunstgeschichte. *Lexicographica* 18: 89-103.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer.
- Tarp, S.** 2009. Reflections on Data Access in Lexicographic Works. Nielsen, S. and S. Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 43-61. Amsterdam/Philadelphia: John Benjamins.
- Tono, Y.** 2009. Pocket Electronic Dictionaries in Japan: User Perspectives. Bergenholtz, H., S. Tarp, and S. Nielsen (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 33-67. Berlin: Peter Lang.
- Trap-Jensen, L.** 2005. Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. *LexicoNordica* 12: 109-121.
- Wiegand, H.E.** 1988. Wörterbuchartikel als Text. Harras, G. (Ed.). 1988. *Das Wörterbuch. Artikel und Verweisstrukturen. Jahrbuch 1987 des Instituts für deutsche Sprache*: 30-120. Sprache der Gegenwart LXXIV. Düsseldorf: Schwann.
- Wiegand, H.E.** 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexicographica* 24: 209-316.

Improving the Functionality of Dictionary Definitions for Lexical Sets: The Role of Definitional Templates, Definitional Consistency, Definitional Coherence and the Incorporation of Lexical Conceptual Models

Piet Swanepoel, *Department of Afrikaans and Theory of Literature, University of South Africa, Pretoria, South Africa (swaneph@unisa.ac.za)*

Abstract: This article focuses on some of the problems raised by Atkins and Rundell's (2008) approach to the design of lexicographic definitions for members of lexical sets. The questions raised are how to define and identify lexical sets, how lexical conceptual models (LCMs) can support definitional consistency and coherence in defining members of lexical sets, and what the ideal content and structure of LCMs could be.

Although similarity of meaning is proposed as the defining feature of lexical sets, similarity of meaning is only one dimension of the broader concept of lexical coherence. The argument is presented that numerous conceptual lexical models (e.g. taxonomies, folk models, frames, etc.) in fact indicate, justify or explain how lexical items cohere (and thus form sets). In support of Fillmore's (2003) suggestion that definitions of the lexical items of cohering sets should be linked to such explanatory models, additional functionally-orientated arguments are presented for the incorporation of conceptual lexical models in electronic monolingual learners' dictionaries. Numerous resources exist to support the design of LCMs which can improve the functionality of definitions of members of lexical sets. A few examples are discussed of how such resources can be used to design functionally justified LCMs.

Keywords: DEFINITIONAL TEMPLATES, DEFINITIONAL CONSISTENCY, DEFINITIONAL COHERENCE, LEXICAL CONCEPTUAL MODELS

Opsomming: **Verbetering van die funksionaliteit van woordeboekdefinisies vir leksikale versamelings: Die rol van definisiematryse, definisie-eenvormigheid, definisiesamehang en die inkorporering van leksikale konseptuele modelle.** Hierdie artikel fokus op sommige van die probleme wat ter sprake kom deur Atkins en Rundell (2008) se benadering tot die ontwerp van leksikografiese definisies vir lede van leksikale versamelings. Die vrae wat gestel word, is hoe leksikale versamelings gedefinieer en geïdentifiseer moet word, hoe leksikale konseptuele modelle (LKM's) definisie-eenvormigheid en

-samehang kan ondersteun by die definiëring van lede van leksikale versamelings en wat die ideale inhoud en struktuur van LKM's sou kon wees.

Alhoewel betekenisoreenkoms as dié definiërende kenmerk van leksikale versamelings voorgestel word, is betekenisoreenkoms net een dimensie van die breër konsep van leksikale samehang. Die argument word aangevoer dat verskeie konseptuele leksikale modelle (bv. taksonomieë, lekemodelle, rame, ens.) in werklikheid aandui, motiveer of verduidelik hoe leksikale items saamhang (en dus versamelings vorm). Ter ondersteuning van Fillmore (2003) se voorstel dat definisies van die leksikale items van samehangende versamelings met sulke verduidelikende modelle gekoppel moet word, word bykomende funksioneel-georiënteerde argumente aangebied vir die inkorporering van konseptueel leksikale modelle in elektroniese eentalige aanleerderswoordeboeke. Talle bronne bestaan vir die ondersteuning van die ontwerp van LKM's wat die funksionaliteit van definisies van lede van leksikale versamelings kan verbeter. Enkele voorbeelde word bespreek van hoe sulke bronne aangewend kan word vir die ontwerp van funksioneel gemotiveerde LKM's.

Sleutelwoorde: DEFINISIEMATRYSE, DEFINISIE-EENVORMIGHEID, DEFINISIESAMEHANG, LEKSIKALE KONSEPTUELE MODELLE

1. Introduction

In arguing for the use of templates to define the meaning of words belonging to the same lexical set, Atkins and Rundell (2008: 124) provide the following two definitions from a learners' dictionary:

- (1) **lion** ... a large strong African and Indian animal with four legs and light brown fur which eats meat and belongs to the cat family.
- (2) **tiger** ... a large wild cat which has yellowish orange fur with black stripes.

The problem is, as they note, that both definitions refer to the animals' size, their fur, and their membership of the cat family. However, the definition for *lion* also provides information about diet, strength, number of legs, and habitat, which would be equally relevant features for the definition of *tiger*. This problem of inconsistency in defining members of lexical sets can be addressed, as the authors propose, by designing a definitional template, such as those in (3) and (4), as part of a larger template entry for the whole lexical set. Lexicographers can then design definitions for the cat/animal lexical set for a specific dictionary using the stipulated attributes/values as listed in the definitional templates in (3) and (4), bearing in mind the dictates of dictionary type, intended users and dictionary functions (Atkins and Rundell 2008: 125).

- (3) LU meaning:
Domain = zoology
The specific animal (species)

a [size] [wild/domesticated] [carnivorous/herbivorous] mammal, Latin name XXX, having fur/hide [colour, markings], found in [habitat]. Also called XXX.

- (4) LU meaning
Domain = zoology
The genus e.g. the cat family, the big cats
a [size] [wild/domesticated] [carnivorous/herbivorous] mammal, of the genus (LATIN NAME), such as the (SPECIES NAME), having fur/hide [colour, markings], found in [habitat]

In addition to the problem of definitional (in)consistency throughout a dictionary, definitions (1) and (2) illustrate another typical problem of many dictionaries, one which I will designate *definitional (in)coherence*. Definitional incoherence takes on a number of forms, but in (1) and (2) it is evident from the fact that *lion* is clearly indicated as being a member of the *animal* lexical set, while *tiger* is not. In this case, *tiger* is only identified as a member of the lexical set of *large wild cats* — a subdomain of the *animal* lexical set. Furthermore, the cohyponyms of *tiger* and *lion* are not mentioned in either of the definitions — not even the fact that they are cohyponyms of the same lexical set (i.e. of large cats). Apart from definitional inconsistency, there is thus also the problem of definitional incoherence in so far as it is not indicated what the members of lexical sets are and how they cohere (or in fact can be considered a lexical set).¹ Lexical coherence is generally taken as a key organizing principle of the lexica of natural languages, and, as will be discussed below, needs to be addressed in dictionary definitions if they aim to support the functionality of such dictionaries.

Practical as the approach of Atkins and Rundell (2008) may be, these two problems are not, or not adequately, discussed by them, and some of the guidelines they provide for addressing these problems give rise to a number of questions that need further consideration.

The first of these is how the concept of "lexical set" should be defined and how lexical sets are to be identified in practice. Atkins and Rundell (2008: 128) mention that nearly sixty lexical sets were identified in a recent dictionary project which were amenable to template entry treatment, but they do not indicate how these lexical sets were derived and what the criteria for set membership was or how these criteria themselves were derived. This leads to the following question:

- (a) How can you define and identify a lexical set in a language? (definition of the object; methodology of identification)

The second problematic aspect is the assumption that definitional consistency (in the choice of definitional features and their structuring) for the lexical items in a set results in more functional definitions, keeping in mind that functionality in dictionary terms is a multidimensional concept (cf. the discussion in Sec-

tion 3). Although a lexical template such as (3) provides a basis for the construction of definitions for specific users, dictionary types and functions, it is an empirical question whether or not definitional consistency for the lexical items in a set will indeed improve the functionality of these definitions. For example, does one have to use specific attributes/values from template (3) for each of the possible functions that a learners' dictionary has to serve or will certain choices have an impact on the functionality of the definitions and others not? A simple example: Would the shorter definition of *tiger* (as in definition (2)) necessarily be less comprehensible and, for example, not assist a user in the interpretation of the meaning of the word in a text (two dimensions of functionality) than would an expanded definition which incorporates all the features listed in the definitional templates (3) and (4)? Whether or not definitional consistency in fact improves the functional quality of definitions is an empirical one, and one for which Atkins and Rundell (2008) provide no empirical backing by way of theory-driven empirical research. This leads to the following research question:

- (b) Does definitional consistency improve the functional quality of dictionary definitions?

The next set of questions relate to the identification of lexical sets, the choice of defining attributes/values for a set, the structuring of such attributes/values in definitions, and the kind of justification one could provide for such choices:

- (c) How do you decide what attributes and features are relevant for the definitions of the lexical items of a specific lexical set? (the question of methodology)
- (d) Of all the possible attributes/values that could be used in the definitions of the lexical items in a set, which of these could/should you choose and on what do you base this decision? (the question of choice of content and justification for this choice)
- (e) How do you structure the chosen definitional features, i.e. in what order do you present the attributes/values in the definition and why? (the question of structure)
- (f) In what form do you present the defining attributes/values, i.e. what language do you choose to encode the attributes/values in the definition? (the question of form) (Cf. Atkins and Rundell 2008: 431-439 for an in-depth discussion of this issue.)

Limitations of space do not allow an in-depth discussion of all the issues raised by these questions. In Section 2 of this article, the focus falls on some of the problems raised by questions (a) and (c)–(e) and some of the solutions proposed by Atkins and Rundell (2008). As will be argued, although similarity of meaning is proposed as the defining feature of lexical sets, similarity of meaning is only one dimension of the broader concept of lexical coherence. The

argument is raised that numerous conceptual lexical models (e.g. taxonomies, folk models, frames, etc.) (abbreviated to: LCMs) in fact indicate, justify or explain how lexical items cohere (and thus form sets). Following Fillmore's (2003) argument that the lexical items of cohering sets should be linked to such explanatory models, additional functionally-orientated arguments are presented in Section 3 for the incorporation of conceptual lexical models in electronic monolingual learners' dictionaries. In electronic dictionaries or web-based dictionaries, such linking mechanisms in definitions allow for the "spreading" of the required linguistic/conceptual knowledge required to understand, use or learn the meaning of members of lexical sets.

Given the fact that conceptual lexical models come in all shapes and sizes and often in human unfriendly formats, they have to be adjusted for human use. Section 4 provides an analysis of one such model.

2 Lexical sets and their design features

2.1 Defining *lexical set*

The problem of defining the term *lexical set*, like its often used synonym *lexical domain*, has a long history in lexical semantics (cf. Faber and Mairal Usón 1999; and Murphy and Medin 1985). Atkins and Rundell (2008: 124) define a lexical set as any "group of words that share a common element of meaning such as days of the week, or months of the year, or birds, trees, flowers, and metals". What the authors most probably had in mind — but do not state clearly — are lexical sets based on the numerous sense relations distinguished, for example, in the WordNet project (synonymy, hyponymy, or troponymy, etc; cf. Baker, Fillmore and Lowe 1998; and Miller and Fellbaum 1991) or belonging to the same semantic domain/field (e.g., the set of communication verbs: *ask, cite, explain, show, tell, babble, bark, bawl, bellow, speak, talk, argue, chat, announce*, etc.; cf. Baker and Ruppenhofer 2002 for a discussion).

However, the authors do not explain what constitutes a legitimate "common element of meaning" to determine whether a specific selection of lexical items constitutes a set (i.e. what attribute(s)/value(s) are to be taken as critical for the similarity metric). Given that no restrictions are imposed on the concept of "semantic similarity" nearly any and all words in a lexicon can form a set with any other word. The saying, "similarity lies in the eyes of the beholder", is reiterated as follows by Murphy and Medin (1985: 292):

Any two entities can be arbitrarily similar or dissimilar by changing the criterion of what counts as a relevant attribute. Unless one can specify such criteria, then the claim that categorization (i.e. of belonging to the same lexical set — PHS) is based on attribute matching is almost entirely vacuous.

As Faber and Mairal Usón (1999) argue, linguists have seldom tried to define the term *semantic domain* in a precise and clearly restricted manner, and the

same goes for what constitutes a lexical set. Faber and Mairal Usón (1999) do, however, provide a clear method for determining the lexical sets in a language by a systematic bottom-up analysis of the semantic attributes/features used in the dictionary definitions of the lexical items of a language. This method at least provides one with a *justified* similarity metric for determining the lexical sets of a language.

The problem of providing adequate definitions of *lexical set* and *semantic domain* will not be explored here any further. As Faber and Mairal Usón (1999: 79) note, the concepts of semantic fields, lexical domains and thus also lexical sets, have been and still is the object of much imprecision. Of more importance is the fact that the problem of the similarity metric for set membership ties in with that of definitional (in)coherence in as much as "similarity of meaning" is but just one of the many ways in which the lexical items of a set may in fact cohere. Both Faber and Mairal Usón (1999) and Murphy and Medin (1985) therefore put forward the case that lexical coherence is a more apt way of defining lexical sets and that in various linguistic theoretical frameworks a number of constructs (other than the kind of taxonomy underlying the *animal set*) are used to explain how individual lexical items cohere and therefore constitute a lexical set.

For such a descriptive, often explanatory, model the term *lexical conceptual model* (LCM) is used in the rest of the article. As used here, LCM is equally general as Kövecses' (2006: 369) term *frame* (derived from the theoretical framework of frame semantics of Fillmore (cf. Fillmore 2003; and Fillmore and Atkins 1992)), which he defines as follows and links to the various other terms used for such cognitive constructs:

Frames are structured mental representations of an area of human experience (i.e., objects or events) ... Framelike structures have received a variety of names in the literature, including model, idealized cognitive model, domain, script, scene, experiential gestalt, folk theory, and several others. Frames have roles ... that can be instantiated by particular values ... Frames are like schemas in that they sanction more specific instances.

To the list of LCMs identified by Kövecses (2006), one could add *folk model*, *cultural model*, *expert model*, *ontology*, *taxonomy* and *hierarchy*. To these Murphy and Medin (1985) add *prototype categories* and *radial categories* (given that deviants from the prototype or prototypical exemplars are also lexicalized). All of these can provide, although they do so in different ways, an analysis/description of the meaning of the members of lexical sets and the nature of their coherence. Obviously, this goes further than "semantic overlap", but includes coherence based on, for example, causality, transformation, contiguity, etc. Fillmore's frames, for example, are defined on the FrameNet website (<http://framenet.icsi.berkeley.edu>) as "schematic representations of situation types (eating, spying, removing, classifying, etc.) together with lists of the kinds of participants, props, and other conceptual roles that are seen as components of such situa-

tions". In as much as the participants, props, etc. of such event types are lexicalized (e.g. *grocer*, *client*, *groceries*, *credit card*, etc.) in a language, the event structure/type (e.g. buying groceries in a store) of the frame provides an explanation of why such words cohere, i.e. they refer to the same frame but have no semantic overlap. (Cf. Murphy and Medin 1985 for an in-depth discussion of how such models explain (as theories about the world) the coherence of members of lexical sets.)

Although the claim is often made that such models represent language users' lexical knowledge, I would like to include under the category of LCMs any systematic description and analysis of the semantic coherence of lexical sets, such as, for example, a linguistically justified analysis of the semantic features of emotion words or of motion verbs and the way in which they cohere, but for which no claims as to their psychological reality are necessarily made (cf. the examples discussed in Section 4).

2.2 Identifying lexical sets

As noted, Atkins and Rundell (2008: 128) provide no clear guidelines for the identification of lexical sets. The imprecision with which lexical sets are defined, no doubt adds to this problem. However, once one starts looking at the work being done on semantic fields, domains, sets, ontologies and frames in numerous disciplines, it becomes evident that there is an abundance of resources which lexicographers can tap into (and sometimes also derive their defining attributes/values from):

- analyses of ontological domains (cf., e.g., Jiménez Briones 2007; Jiménez Briones and De Alba 2008; Keil 1979, 1981; Vinson et al. 2003; Vinson and Vigliocco 2002; and Pustejovsky et al. 2006),
- systematic analyses of semantic and syntactic similarities and differences between members of a lexical class and between subclasses of a larger lexical class (cf., e.g., De Clerck et al.'s (To appear) analysis of the verbs of instrument of communication; and Moreno's 2007 analysis of induced motion verbs),
- analyses of the attributes/values used in dictionary definitions and the generation of lexical sets from such analyses "from the bottom up" (cf. e.g., parsing of the attributes/values of the definitions in (machine readable) dictionary texts in Calzolari 1992; Jiménez Briones 2007; Jiménez Briones and De Alba 2008; and De Boni and Manandhar 2002),
- thesauri, such as *Roget's Thesaurus* (1987),
- large structured electronic dictionaries such as FrameNet and WordNet (cf. Fellbaum 1990; Miller 1990; and Miller and Fellbaum 1991), sometimes enriched with other specialist theories of the mental lexicon (e.g.

the generative lexicon of Pustejovsky (cf. Pustejovsky 1990) as used in, for example, Liu and Wu 2003 (cf. also De Boni and Manandhar 2002 on how to enrich WordNet with telic information), and

- cultural folk models of a certain knowledge area (e.g. ethnobiological nomenclature, the culture-specific model of the concept SELF which explains the meaning of a large number of *self* compounds in English (cf. Morillas 1997); and the folk theory of the mind (cf., e.g., D'Andrade 1987; Malle 2005; and Keesing 1987).

Examples are the following:

In *Roget's Thesaurus* six main conceptual categories are distinguished, each of which denotes a large lexical set and which could be further subdivided into smaller lexical sets: (a) Abstract relations, (b) Space, (c) Matter, (d) Intellect: the exercise of the mind, (e) Volition: the exercise of the will, and (f) Emotion, religion and morality.

In a large database such as WordNet 25, unique beginners are discerned which each identify a large semantic domain/lexical set, some of which can be schematized in a typical hyponymy model as in Figure 1 with the beginners denoting different kinds of tangible things.

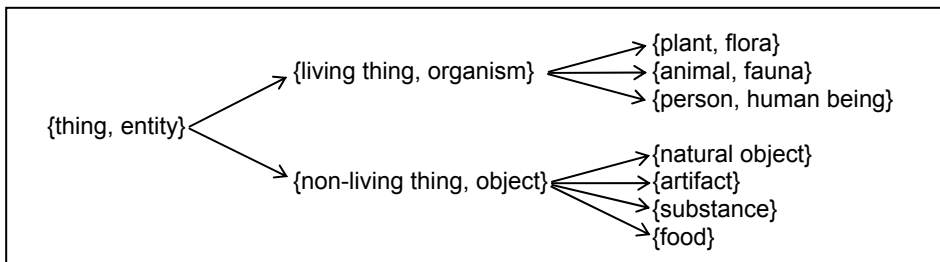


Figure 1: Diagrammatic representation of hypernymic relations among unique beginners denoting different kinds of tangible things (Miller 1990: 252)

In WordNet, verbs are ordered in fifteen semantic domains/lexical sets, for example, verbs of bodily care and functions, change, cognition, communication, emotion, motion, etc. (cf. Fellbaum 1990).

In a more recent contribution, Liu and Wu (2003) show how the fourteen frames in FrameNet pertaining to the domain of communication (*candidness, commitment, conversation, encoding, gesture, manner, noise, questioning, request, response, statement, volubility, hear* and *means*) can be viewed as subsets of a cognitive system based on the Conduit Metaphor — which gives the necessary coherence to the communication frames. The fourteen frames themselves characterize the distinct background information one has to acquire to understand the senses of the individual verbs of communication where each frame is defined in terms of its prototypical "frame" elements.

As Faber and Mairal Usón (1999: 82) note, however, in most thesauri and databases such as WordNet and FrameNet, the semantic fields/lexical sets (as ontologies) are established *a priori* by the compilers so that one has little guarantee that the domains/sets represent inherent cognitive reference points for language users.

Obviously, each of these methods has its own pitfalls in modelling users' lexical knowledge, but they could offer lexicographers valuable assistance in identifying lexical sets.

2.3 Generating defining attributes and values

2.3.1 "Picking the brains of colleagues"

After relevant lexical sets have been identified, the next issue is how one can identify the relevant attributes and values for the definitions of members of lexical sets. Atkins and Rundell (2008: 127-128) spell out what seems to be a relatively simple procedure where each lexicographer working on a dictionary chooses one word from a specific lexical set (words that are representative of or prototypes of subtypes of a category, e.g. with regard to the animal set *lion, unicorn, rat, cow, cat* and *fox*) and then has to compile the richest corpus-based entry they can for that word. In the next step, these entries are compared and collated in a discussion session and a final version of the template entry is drawn up with all possible relevant lexical units included (i.e. the full lexical set) (which, I assume, will be noted on the larger lexical template).

As suggested above, compilation of a definition template is not left only to the lexicographers' intuition about what semantic attributes and features would constitute the meaning of the members of such lexical sets, but it is also to be controlled by current usage working bottom-up from whatever corpus of usage is employed in the compilation of the dictionary (cf. Atkins and Rundell 2008: 264-380). Corpus data as such does not, however, provide one with the relevant attributes/values to use in definition templates. As Atkins and Rundell (2008: 311) note, defining provisional word senses for a lexical item in a set (based as they are on the choice of specific attributes/values) are "the subjective, intuitive part" of the exercise (and more so when it comes to lexical sets organized in taxonomies, hierarchies or ontologies). As they (Atkins and Rundell 2008: 307) admit though, corpus-based data on other linguistic aspects of these lexical items in fact only "complement our intuitions about meaning and underpin an analysis which is as objective and 'scientific' as it reasonably can be, given the slippery and dynamic nature of word meaning".²

No one would probably disagree with the latter part of this quote — it is no easy task formulating functional definitions. But the authors' views also underline that there is no simple way in which one can harvest lexical sets and their defining features for definition templates (such as (3) and (4)) from corpus data (unless the corpus itself consists of definitions of such lexical sets). How-

ever, there are a number of problems with this approach. Firstly, one cannot assume that a few lexicographers will have direct access to what could/should be taken as the defining features of all possible lexical sets that have to be defined in explanatory dictionaries. Exhaustive lexical knowledge of this kind simply does not reside in a few language users/lexicographers.

Secondly, it remains an open question whether or not there are alternative methodologies available which would also pass the test of being "objective and scientific", however leniently one may use these labels. Identifying lexical sets and their relevant definition attributes/values is a task not only required in lexicography but also in numerous related disciplines which use various kinds of methodologies (equally time-consuming as corpus analysis), such as linguistics, computational linguistics, natural language processing, information design, psycholinguistics, and anthropology. One of these methods is to elicit defining attributes and values for members of lexical sets from language users and then employing clustering techniques to identify the major defining ones (as used, for example, in experimental psychology; cf., e.g., Barsalou 1992; McRae, De Sa and Seidenberg 1997; Vinson et al. 2003; Vinson and Vigliocco 2002; Wu and Barsalou 2009; and Santos et al. To appear).

One of the problems concerning the feature generation methodology, however, becomes apparent if one considers the analysis and coding system for generated features for nouns devised by Barsalou and colleagues (cf. Wu and Barsalou 2009; and Santos et al. To appear). As they indicate, language users can generate defining features for nouns which denote any aspect of a very complex conceptual system associated with the members of lexical sets — something perhaps broader in conceptualization than Fillmore's (2003) frames (cf. Barsalou 2003; Barsalou and Wiemer-Hastings 2005; Wu and Barsalou 2009; and Santos et al. To appear). In the most recent edition of their coding system (personal correspondence Barsalou), the following are listed as relevant feature sets to describe the relationships between a target concept and the features that could be generated in a feature generation exercise. Such a list makes clear what kinds of definitional attributes language users could in principle generate for defining members of lexical sets.

- Taxonomic categories (synonymy, ontological category, superordinate, coordinate, subordinate, individual value of an attribute)
- Situation properties (person, living thing, object, social organization, social artifact, building, spatial relation, time, action, event, manner, function, physical state, social state, quantity)
- Introspective properties (affect/emotion, evaluation, representational state, cognitive operation, contingency, negation (absence of something), quantity)

The question is: Which of these features should be used in designing functional definitions for members of lexical sets, or should all of them be used, given that

they constitute the elements of the meaning of a lexical unit? On the other hand, many of these possible features for lexical sets are already accommodated in dictionary definitions: relational attributes (synonymy, hyponymy, troponymy, etc.) and decompositional attributes (cf. Miller and Fellbaum 1991 for a summary). In as much as situational attributes overlap with frame³ information, Fillmore (2003) presents a proposal as to how such frame information could be incorporated into dictionary definitions (cf. the discussion in Section 3).

2.3.2 Using genus and differentia definitions

Another way to approach the question of what defining attributes/values are to be used in definitions of members of lexical sets is to use the traditional genus and differentia definition. Atkins and Rundell (2008: 415) in fact view the traditional genus and differentia definition as one of the most useful ones for defining the members of various kinds of lexical sets. Typically this kind of definition type consists of a superordinate term (genus) which places the lexical item in a specific semantic category (lexical set) and additional information (differentia) which indicates what makes the lexical item under consideration unique and in what ways it differs from other members of the same category (its cohyponyms). This definition type is often used not only for certain noun classes, but also for verb classes. The authors illustrate this type of definition with the following example from ODE-2 2003:

- (5) **convertible** ... car (*genus*) with a folding or detachable roof (*differentia*)

where the differentiating feature (*with a folding or detachable roof*) distinguishes *convertible* from its cohyponyms *saloon*, *estate car* or *people carrier* (cf. Atkins and Rundell 2008: 414).

However, there are problems with this kind of definition and with its use in defining members of lexical sets. It presupposes that dictionary users know what the superordinate/genus *car* means as this forms part of the meaning of *convertible* (but is not spelled out), which in turn presupposes that users also have knowledge of the larger hyponymy relation which underlies this definition, viz. convertible → car → (vehicle → ... artifact). This follows from the fact that each superordinate term inherits as hyponym some of its defining features from its own superordinate term. In the case of *car*, one could assume that most dictionary users would possess such knowledge; however, in other taxonomies this may not be the case. In short: The genus only places a lexical item in a semantic category; it does not define the meaning of that category so that the relevant attributes/values in terms of which the lexical items of the category are to be defined are made explicit.

In a more or less "behind the scene" manner, Atkins and Rundell (2008: 418) justify the choice of the defining attributes and values in the ODE-2 2003

definition of *bus* by indicating how each of the differentia in fact differentiates a bus from other types of motor vehicle:

- (6) **bus** ... a large motor vehicle carrying passengers by road, typically one serving the public on a fixed route and for a fare.

Their explanation is as follows (Atkins and Rundell 2008: 418):

- a bus is large (and so it is not a car or a taxi)
- it goes by road (so it's not a train)
- it carries passengers (so it's not a truck)
- it typically operates on a fixed route and charges its passengers a fare (which are the features of the prototypical bus and which differentiate it from other bus types that deviate from these prototypical features: school buses, hotel buses, airline buses, staff buses, etc.

The problem is, of course, that the explanatory information, i.e. the explanation for the choice of defining features and the taxonomic structure of the category of vehicles which it refers to, is not typically provided for the dictionary user, although having access to it becomes crucial in acquiring the relevant concept (cf. the discussion in Section 3.1.2).

Thus, it is not always clear precisely by what cohyponyms (and their specific semantic features) a word is differentiated if the rest of the lexical set is not given with a clear indication what their (semantic) features are and precisely how they differ from each other. For example, none of the cohyponyms of *convertible* or those of the genus *car* is provided in the definition so that the user in fact does not know from what other types of cars a convertible is distinguished.

These problems also become evident in (1) and (2) and through the definitional templates (3) and (4) (cf. Section 1) in which the attributes/values refer to a larger cognitive structure, namely an ethnobiological animal taxonomy in the case of (1) and (2) and a scientific taxonomy in the case of (3) and (4). In the case of (3) and (4), the domain name/label *zoology* and the Latin terms of the genus and the species within the definition templates link the lemma to a scientific taxonomy.

A further problem with the genus and differentia definition is that it omits much of the conceptual content/meaning associated with the lexical item that is defined. Examples (1) and (2) are a case in point: The definitions provide the dictionary user with the immediate genus (cat family) and a higher superordinate (animal) plus information that is supposed to differentiate the lexical entries from all others in the lexical set (i.e. they are typical genus and differentia definitions). However, these definitions also omit much information on the semantics of the two lexical items that belong to the same set. Firstly, as already indicated by Atkins and Rundell (2008), they do not provide all the information people have of these animals. The question would therefore be, if someone has

no idea of what either *tiger* or *lion* means, whether or not that person, given these two definitions, will be able to construct accurate concepts of these two lexical items. As Miller (1990: 247) concludes: There is ample support for the inclusion of more, and even so-called "encyclopedic", information into dictionary definitions than merely the genus and one differentiating feature. This approach is evidenced by the NOED definition in (7) (conforming more to the lexical template (3)), which not only provides more information in the main definition, but also incorporates numerous other features in the "encyclopedic" component of the lexical entry.

- (7) **cat** ... a small domesticated carnivorous mammal with soft fur, a short snout, and retractile claws. It is widely kept as a pet or for catching mice, and many breeds have been developed.
- *Felis catus*, family Felidae (the **cat family**); probably domesticated in ancient Egypt from the local race of wild cat, and was held in great reverence there. The cat family also includes the ocelot, serval, margay, lynx, and the big cats ... See also **BIG CAT**.

Neither of the definitions in (1) and (2) contains information about the cohyponyms of the lexical entries which make up the rest of the lexical set. As Miller (1990: 246) notes, dictionary definitions typically point upwards to a superordinate term, but not sideways to coordinate terms or downward to hyponyms. However, this is not general practice as is evidenced in definition (8). In as much as the other information in the definition is supposed to differentiate the lexical headword from other members of the set, it is in fact not clear from which one or more of the rest of the lexical set it is differentiated.

The hierarchies presented in dictionaries are often rather flat (if one considers, for example, that the superordinate structure of a lexical item such as *canary* is six levels deep: canary → finch → passerine → bird → vertebrate → animal; cf. Miller 1990: 250). In as much as the meanings of superordinate terms are inherited by those below them, one is given no indication in the definition of what the superordinate lexical items themselves may mean. Here again, the dictionary user would have to search for the meaning of the superordinate terms, if they are indeed entered in the dictionary. For example: In the NOED, there is no separate entry for *cat family*, but it is simply found as undefined term in the entry for *cat*. Furthermore, there is a reference to *big cat* as a separate entry, but without any indication that a difference is made between the small and the big cats and thus that lion and tiger not only belong to the cat family as indicated in (1) and (2), but more specifically to the family of big cats.

- (8) **big cat** ... any of the large members of the cat family, including the lion, tiger, leopard, jaguar, snow leopard, clouded leopard, cheetah, and the puma.
- *Panthera* and the other genera, family Felidae.

As is the case with this kind of relational information, one has to start searching in the dictionary to find the relevant information, while it could, as will be

argued below, be made available by incorporating a clickable link in the definition of each of the entries for each of the *Felidae* lexical items that transports the user to a large LCM where all lexical items are linked in a taxonomy (with definitions for each of the lexical items).

Atkins and Rundell (2008: 416-417) are well aware of some of the other major problems with genus and differentia definitions in so far as they sometimes lead to definitions which attempt to list every possible defining feature to cover every possible instance of the category to which a lexical unit refers, or at other times they result in short and vague definitions in an attempt not to exclude any members of the category. Indeed, finding the balance between brevity in definition, comprehensibility and coverage is not an easy matter, especially if the lexicographer only has the lexicographic definition at his/her disposal to explain the semantics of (members of) lexical sets. As will be argued below, however, linking dictionary definitions with various kinds of dictionary frames, whether internal and external, offer the opportunity to spread out the relevant semantic information in dictionaries, allowing users freedom to access as much information as possible for whatever purpose they consult an explanatory dictionary.

2.4 The structure of definitions of lexical sets

In addition to the question of what defining attributes/values to select for the definitions of members of lexical sets, there is also the question of the order in which the chosen attributes/values should be presented. Is there some kind of "internal logic" in the order in which the defining attributes/values of definitions of members of lexical sets are given, or would it have no functional consequences if some random order was chosen?

Atkins and Rundell (2008: 439) address this issue of the ordering of the attributes FORM (what a thing looks like, is made of) and FUNCTION (what it is used for) as in the definition of artifacts such as *windmill*. No clear rules for the ordering of the attributes are given, but they present Bolinger's (1965: 572) suggestion that the first attribute chosen should relate "the unknown to the known" (Atkins and Rundell 2008: 439). Consequently they suggest that in a dictionary aimed at users in industrialized countries where windmills are rarely used for their original function of crushing grain, the first attribute for *windmill* should be its form (as windmills are familiar structures with a distinctive form) (cf. (9) for the definition from ODE-2 2003):

- (9) **windmill** ... a building with sails or vanes that turn in the wind (FORM) and generate power to grind corn into flour (FUNCTION).

On the other hand, for the definition of *watering can*, the criterion is how much of the definition one would have to read before being able to identify the definiendum. According to this criterion, they would prioritize definition (11) of *watering can* over definition (10). However, there is in my opinion no way to

choose in this regard between (11) and (12) where the order of the attributes are reversed, but both are, of course, very short definitions.

- (10) **watering can** ... a container with a tube that ends in a wide mouth with many small holes (FORM) which is used for watering things, usually plants (FUNCTION).
- (11) **watering can** ... a container used for pouring water on plants (FUNCTION), with a handle and a long spout (FORM).
- (12) **watering can** ... a container with a handle and a long spout (FORM), used for pouring water on plants (FUNCTION).

As a guideline, Atkins and Rundell (2008: 440) further note, firstly, that if form and function are presented in optimal order, the chances are increased that users may "log off" before the end of the definition, and secondly, that form is a less reliable indicator of the meaning of a lexical item "since the shape and construction of things tend to vary but their function is usually stable".

However, form and function are only two of the many possible attributes that could be used in defining the members of lexical sets. For the description of nouns (specifically artefacts such as windmills and watering cans), a more justified set of definitional attributes is proposed by Pustejovsky for his generative lexicon. Artefacts can, for example, be defined in terms of any one or more of the following attributes:

- Constitutive relationships (the relation between an object and its constitute parts, e.g. material, weight, parts and component elements — attributes which could also be regarded to fall under FORM)
- Formal relationships (that which distinguishes an object within a larger domain, e.g. orientation, magnitude, shape, dimensionality, colour, position)
- Telic relationship (the purpose or function of an object, e.g. the purpose an agent has in performing an act or the built-in function or aim which specifies certain activities — attributes that would fall under FUNCTION)
- Agentive relationships (factors involved in the origin or "bringing about" of an object, e.g. creator, artifact, natural kind, causal chain)

(Cf. also Bougarev and Pustejovsky 1990; De Boni and Manandhar 2002; Calzolari et al. 2005; and Cimiano and Wenderoth 2007.) These attributes play a crucial role in the semantics of the compounds formed with such nouns (as heads) and their polysemy. Once again, there is no strict ordering of the attributes in definitions utilizing the telic structure.

Generally speaking though, language users associate richer conceptual structures with lexical items, and thus with the members of lexical sets. For

example, nouns which fall within the basic level in any hierarchy/taxonomy also denote things with which people have sensory motor interaction (form: perception, taste, touch, smell, sound; function: motor interaction). Language users thus have extensive conceptual knowledge (attributes) of such entities, but there is clearly no set order for such attributes and neither is there one attribute which dominates in all contexts.

As far as could be ascertained, no research has been forthcoming on the optimal ordering of the defining attributes of lexical (sub)sets. One way to proceed would be to work from the definitions for the members of lexical subsets and ascertain what attributes/values are commonly used and what ordering patterns occur before experimentally testing what ordering(s) is (are) optimal for which kinds of users, dictionary types and dictionary functions.

3. Using lexical conceptual models in defining members of lexical sets — functionalist considerations⁴

Most of the problems related to the definition of members of lexical sets discussed above can be addressed by enriching their definitions in electronic dictionaries with linked/clickable LCMs.

Fillmore (2003) provides a convincing argument for the inclusion in electronic monolingual learners' dictionaries of "double-decker" definitions for sets of lexical items that refer to the same LCM, in this case, frame. A double-decker definition contains two parts: one that provides a semantic explanation (definition) of the meaning of the lexical item, and one which provides a (clickable) link to a frame of structured background information which dictionary users could need to understand not only the semantic explanation of the lexical item under consideration but also the explanations/definitions provided for the other lexical items in a lexical set. The following are a few examples:

Most dictionary definitions for the lexical set *id*, *ego*, and *superego* only become transparent if they are interpreted within the frame of Freud's theory of primitive psychic energies and the manner of their control and modification in the maturing individual (Fillmore 2003: 272-275). Most definitions of the lexical items that name the days of the week only make sense (for non-Western dictionary users) if they are defined against the background frame of the Western calendric concept. As Fillmore (2003: 267) notes: "A dictionary definition that identifies *Wednesday* merely as the middle day of the week is sufficient only when the full background (i.e. frame of the calendric concept — PHS) can be taken for granted." Likewise, definitions of *heaven*, *hell*, *purgatory*, and *limbo* can only be understood against the folk theory of Catholic eschatology (cf. Fillmore 2003: 279-283). Such frames have to be provided if, as Fillmore (2003: 284) says, "'outsiders' are to end up having the same understandings as the people who live within these frames".

In general, Fillmore's (2003) frames, like schemas, scripts, image schemas (cf. Fillmore 2003: 288), Idealized Cognitive Models (ICMs) (cf. Lakoff 1987)

and cognitive cultural models (cf. Morillas 1997) are encyclopedic knowledge structures or conceptualizations underlying the meaning of sets of lexical items that in some way appeal to such structures and that need to be accessed (either in your own brain if "you live the frame", or some encyclopedic work if you do not) to make sense of or to understand the dictionary definitions of the lexical items in such sets. By explicating such mental models, learners' dictionaries can thus give dictionary users insight into the semantic coherence which sets of lexical items with the same frame have — something that the traditional, printed, alphabetically organized learners' dictionary for the most part cannot.

Two of the advantages of using such frames are that the LCM provides the attributes/values to be used in definitions of members of a lexical set, and that not all the attributes/values have to be included in the definition of each member of a set. Members of lexical sets are all linked to the same frame. Thus the relevant background information is provided only once and dictionary users can decide whether or not they have to access the frame for more information to comprehend a definition or not. Another advantage is that the user is not obliged to start an extensive look-up exercise as in a printed dictionary to find the relevant information, even if it is included in the same dictionary.

The inclusion of LCMs in electronic learners' dictionaries is also justified by functional considerations. With regard to their functions, Bergenholtz and Tarp (2003) distinguish between the knowledge and communication functions of a dictionary (cf. also Gouws and Prinsloo 2005). Knowledge functions relate to situations where users for one reason or another want to obtain additional information on some topic, e.g. general cultural and encyclopedic information, specialized information regarding a scientific discipline (e.g. biology, geology etc.) or information about a specific language related to the language-learning process (for example the learning of a foreign language).

For the communication functions, dictionaries have to provide users with the necessary data for the following communication tasks:

- text production in the native or in a foreign language,
- text reception in the native or in a foreign language, and
- translation of texts from the native language to a foreign language and vice versa.

An adequate understanding of the meaning of members of lexical sets is a prerequisite for a dictionary to be able to assist the user in any of the knowledge and communication functions listed above. In so far as linked LCMs can assist in this regard, their inclusion is completely justified.

Fillmore (2003) refers to such definitions as "lexico-encyclopedic" definitions. As the term clearly suggests, Fillmore, as is the case in most cognitive semantic theories, makes no principled distinction between so-called linguistic and encyclopedic lexical meaning. Likewise, dictionaries linked with LCMs can be considered lexical knowledge databases instead of mere lexical/lexico-

graphic databases. Such additions are necessary to support the multiple functions of learner's dictionaries.

3.1 Providing information for the knowledge functions

3.1.1 Cultural and encyclopedic information, specialized information regarding a scientific discipline

The first argument for the inclusion of LCMs in dictionaries is provided by the knowledge function. This requires that dictionaries should provide users with general cultural and encyclopedic information, specialized information regarding a scientific discipline (biology, geology etc.) or information about a specific language related to the language-learning process (for example the learning of a foreign language).

Although Bergenholtz and Tarp (2003) most probably had in mind that these three categories of information should be provided in different kinds of dictionaries or lexicographic resources (i.e. encyclopedias, subject dictionaries and linguistic dictionaries), information from the first two traditional resources (encyclopedias and subject dictionaries) may be required as linked LCMs in traditional dictionaries, as the examples discussed in Section 1 and Section 2 clearly indicate.

In as much as specific cultural models, encyclopedic and subject-specific information explain the meaning of specific lexical items or terms which have found their way into the general vocabulary, the definitions of such lexical items will have to be supplemented with their associated cultural, encyclopedic or subject LCMs. Fillmore's (2003) example of how the lexical set consisting of *id*, *ego* and *superego* requires access to an LCM in which Freud's theory of the human psyche is explicated is an example of such a subject-specific LCM required as link in a linguistic dictionary. Any ethnobiological model of fauna and flora, on the other hand, are examples of folk models which explicate the meaning of the large lexical sets denoting plants and animals.

Cultural specific models also explain the meaning of large lexical sets. One example includes the Western/American folk model of mind underlying the meaning of lexical sets related to the processes of thinking, intention, desires and emotion (cf., e.g., D'Andrade 1987; and Keesing 1987). Another example is the very complex LCM of the self which underlies an extensive set of lexical items, especially compounds with *self*- in English (cf. Morillas 1997). Many common nouns and verbs, even some proper nouns, which are headwords in encyclopedias also appear in monolingual dictionaries, and, once a strict distinction between linguistic and encyclopedic information is no longer kept, linking the two by way of lexicographic definitions and LCMs is a natural second step.

3.1.2 Acquisition of a foreign language

In providing information for users for acquiring a foreign language, the lexicographer specifically has the task (in designing the monolingual explanatory dictionary), as Fillmore (2003: 284) notes,

to make explicit the background of beliefs, experiences, practices, institutions, or ready-made conceptualizations available to the speakers of the language as the necessary underpinnings of the way they speak and the ways they "think for speaking".

One of the important aspects of learning a foreign language is learning such lexical sets. There is ample psycholinguistic evidence that many of these sets are relationally structured (cf., e.g., Fellbaum 1990; Miller 1990; and Miller and Fellbaum 1991). For example, as in the case of the animal words discussed in Section 1, many lexical sets in any one language are linked via hyponymy relationships. The acquisition of such hyponymy relations in a language are an important aspect of the learning of the lexicon of such a language (cf. Crossley, Salsbury and McNamara 2009). Building such hyponymy lexical relations is an essential aspect of vocabulary acquisition in a language either as one's mother tongue or as a foreign language (cf. Li 2009: 636; however, also see Nation 2000). Linking the members of lexical sets to such taxonomies or hypernymic models is therefore a prerequisite in monolingual explanatory dictionaries which aim to satisfy the (lexical semantic) information needs of language learners.

The lexical sets associated with Fillmore's type of frames are also often used as the basis for constructing learning materials for L2 lexical acquisition, for example, going to a restaurant, the market, going shopping, visiting particular sights, etc. A step further is the production of bilingual dictionaries based on such frames. One such dictionary in the making is the English–Hungarian dictionary of Kövecses (personal communication) with entries such as the following from the musical concert frame. See example (13):

- (13) **karmester conductor** A zenekart a karmester vezényli. The orchestra is conducted/directed by the conductor. FOGALOM: A zenekar (**orchestra**) különböző hangszereken (**instruments**) játszó zenészekből (**musicians**) áll. A zenekart a karmester (**conductor**) vezényli (**conduct, direct**). A zenekar először hangol (**tune (sg) up**), majd a karmester utasítására játszani (**play**) kezd. A karmester a hangmesteri pálca (**baton**) segítségével vezényel a dobogóról (**rostrum, podium**).

Of course, language learners need more information on the conceptual structure of members of lexical sets to acquire these lexical items and to be able to use them in production tasks. Including LCMs such as those discussed in 3.1.1 and 3.1.2 in electronic dictionaries will therefore also support the language acquisition function.

3.2 Providing information for the communication functions

What kind of LCMs is required for comprehension, production and translation? Because of the limitations of space, the first and last of these will be discussed.

Elman (2009) provides numerous examples of how language users' knowledge of event structures (equivalent to Fillmore's frames)⁵ are activated in the processing/comprehension of texts and how this knowledge is used to predict the meaning of the rest of a sentence and to disambiguate polysemous words/sentence fragments. For example, *surgeon* in the sentence fragment *The surgeon ...* activates the hospital-operation LCM/frame, where such an LCM would then list all the lexical items (and their associated meanings) which denote the agents, location, procedures, etc., involving the prototypical event of doing an operation in a hospital. Such an LCM can be primed by any of the members of the lexical set (e.g. *operating table*) and such LCMs, once activated, make it possible to access knowledge of the rest of the LCM, making it unnecessary to repeat much of the information contained in the LCM in normal discourse. However, without knowledge of such an LCM, of which not all relevant conceptual content is provided in a prototypical lexicographic definition, one would make very little sense of what people are talking about. As Elman (2009: 572) notes: "Events play a major role in organizing our experience. Event knowledge is used to drive inference, to access memory, and affects the categories we construct."⁶ (Cf. also Shipley and Zacks 2008 on the role of event knowledge in perception, action and cognition.)

The functionalist considerations given above in fact just add to some of the criteria Atkins and Rundell (2008: 412-413) spell out as good definitional practice, i.e. definitions

- must be intelligible (which would include, among others, that a user should not have to consult another definition to understand the one being looked up),
- should supply enough information to enable the user to understand the word in the context it is encountered,
- enable the user to interpret the word in any new context (so that it enters the user's passive vocabulary), and
- enable the user to use the word correctly in a new context (so that it enters the user's active vocabulary).

4. Designing LCMs

How should the proposed LCMs look like in order for them to be linked to members of lexical sets? In most cases, such as the FrameNet and WordNet ontologies and those devised for use in natural language processing applica-

tions, they contain much of the information needed for the design of LCMs. However, the presentation of the data is usually not user-friendly and often not easily comprehended given the often arbitrary symbols used or the format chosen to encode attributes, values, relations, etc.

On the other hand, other analyses of lexical sets that avoid these problems are more directly usable. One such example is the lexical templates designed as part of the Lexical Construction Model (cf. <http://www.lexicom.es>) for the category of *anger* verbs (Jiménez Briones 2007) and *happiness* verbs (Jiménez Briones and De Alba 2008). Both the English and Spanish lexicons of LEXICOM are organized paradigmatically and syntagmatically into a series of coherent semantic classes/lexical sets (e.g. EXISTENCE CHANGE, POSSESSION, SPEECH, EMOTION, etc.) which are derived by means of a process of exhaustive semantic factorization, working upwards from various dictionary entries (such as those of LDOCE, CIDE and COBUILD) of the possible predicates belonging to each class (for example 250 feeling verbs have been identified, belonging to various subclasses). For every verb a genus (the nuclear meaning) and differentia (adverbial modification (e.g. Manner, Reason, Degree) or other distinguishing features (e.g. selectional restrictions), pragmatic or register features (formal, literary, emphatic) are provided. Furthermore, definitions are encoded in Wierzbicka's Natural Language Semantics, making the definitions themselves very accessible, although its limited nature often does not allow one to express the differences in meaning between members of lexical sets all too clearly.

The usability of these analyses to function as LCMs for human dictionaries is clear from Jiménez Briones's (2007: 2) presentation of *anger* verbs (cf. Table 1).

to cause somebody to feel emotional aversion [anger]

1. **anger**: to cause somebody to **feel** anger.
- 1.1. **annoy**: to **anger** somebody a little.
 - 1.1.1 **vex**: to **annoy** somebody, causing them to feel puzzled. [Old-fashioned].
 - 1.1.2 **displease**: **annoy** somebody, causing them to feel displeasure/inconvenience [Fml.].
 - 1.1.3 **irk**: to **annoy** somebody, causing them to feel disgust [Infml.].
 - 1.1.4 **nettle**: to **annoy** somebody for only a short time.
 - 1.1.5 **gall**: to **annoy** somebody by disappointing them.
 - 1.1.6 **bug**: to **annoy** somebody so that they cannot stop thinking about it [Infml.].
 - 1.1.7 **irritate**: to **annoy** somebody, causing them to feel irritated.
- 1.2. **provoke**: to **anger** somebody deliberately by trying to make them act aggressively.
 - 1.2.1 **needle**: to **provoke** somebody deliberately by repeated criticism [Infml.].
- 1.3. **antagonize**: to **anger** somebody by making them feel hostile towards you.
- 1.4. **rile**: to **anger** somebody very much [Infml.].
- 1.5. **exasperate**: to **anger** somebody very much, causing them to become impatient or frustrated.
- 1.6. **outrage**: to **anger** somebody extremely, causing them to feel offended or shocked.
- 1.7. **incense**: to **anger** somebody extremely, causing them to feel indignation.

- 1.8. **enrage**: to **anger** somebody extremely, causing them to lose self-control.
- 1.9. **infuriate**: to **anger** somebody extremely, causing them to be furious.
- 1.10. **madden**: to **anger** somebody extremely, causing them to figuratively become mad.

Table 1: Paradigmatic organization of *anger* verbs (Jiménez Briones 2007: 2)

All members of the lexical set of *anger* words can thus be linked to this LCM which would make it clear to the dictionary user precisely what lexical items make up the set, and how they share attributes and in what ways they differ from the rest of the members of the set. (Cf. also Faber and Mairal Usón's 1999: 109-115 analysis of the semantics of the manner of walking verbs and Jiménez Briones' 2007 analysis of verbs of feeling.)

5. Conclusion

As Fillmore (2003) notes, technological advances have made it possible to construct lexical databases in which various kinds of text types (dictionaries, encyclopedias, etc.) can be incorporated and linked in numerous ways to enhance the functionality of such reference works. However, he stresses that there is still an urgent need to develop theories and models of lexical conceptual knowledge to support the design of such databases. The development of Frame Semantics and the proposal on how to link frames to the lexicographic description of lexical sets in electronic dictionaries is an important step in this direction. The proposal to link definitions of lexical sets with a larger category of lexical conceptual models is an attempt to take Fillmore's (2003) a small step further. Obviously, much more research is needed on a number of topics, such as the definition of lexical sets, how to generate their definitional attributes and values, and the best designs for LCMs to optimize their functionality in dictionaries of different types and for different kinds of dictionary users.

Endnotes

1. This is not to say that these two problems occur in all dictionaries and to the same degree (cf., e.g., the dictionary entry (8)).
2. Hanks (2000: 211) sums it up as follows: "There is no direct route from the corpus to the meaning. Corpus linguists sometimes speak as if interpretations spring fully fledged, untouched by the human hand, from the corpus. They don't. The corpus contains traces of meaning events; the dictionary contains lists of meaning potentials."
3. On the FrameNet website (<http://framenet.icsi.berkeley.edu>) frames are defined as: "Schematic representations of situation types (eating, spying, removing, classifying, etc.) together with lists of the kinds of participants, props, and other conceptual roles that are seen as components of such situations. The semantic arguments of a predicating word correspond to what we call the frame elements of the frame associated with that word."
4. Although much has been made in the literature about the fact that the design of dictionaries should be determined by their intended users and their major functions, very little research

has in fact been forthcoming in which the results of theory-driven empirical research on these functions are systematically and comprehensively linked with each and every design feature of the various kinds of dictionary types and dictionary users.

5. Elman (2009: 572) defines an event as "a set of participants, activities, and outcomes that are bound together by causal interrelatedness".
6. Given the critical role event knowledge plays in comprehension, Elman (2009) argues, ironically, for a grammar without a lexicon, or at most one in which lexical entries consist only of a lexical headword with a pointer to the event frame that explains the meaning and use of the lexical headword. The same line of argument for a minimalist lexicon is followed in Carter (1997). In Barsalou's semantics of situatedness, a network architecture for lexical knowledge is proposed in which linguistic forms such as lexical items constitute associative networks, but in which these forms are linked to all the defining features of event-like structures.

References

Dictionaries

- CIDE:** *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.
COBUILD: *Collins COBUILD English Dictionary*. London: HarperCollins.
LDOCE: *Longman Dictionary of Contemporary English*. Harlow: Pearson Education.
NOED: *The New Oxford English Dictionary*. Oxford: Oxford University Press.
ODE-2: *Oxford Dictionary of English*. Second Edition. Oxford: Oxford University Press.
Rogets Thesaurus: *Rogets's Thesaurus of English Words and Phrases*. Harmondsworth: Penguin Books.

Other sources

- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Baker, C., C. Fillmore and J. Lowe.** 1998. The Berkeley FrameNet Project. *Proceedings of the COLING-ACL, Montreal, Canada*. http://framenet.icsi.berkeley.edu/index.php?option=com_contentandtask=viewandid=39andItemid=42.
- Baker, C. and J. Ruppenhofer.** 2002. FrameNet's Frame vs. Levin's Verb Classes. Larson, J. and M. Paster (Eds.). 2002. *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*: 27-38. http://framenet.icsi.berkeley.edu/index.php?option=com_contentandtask=viewandid=39andItemid=42.
- Barsalou, L.** 1992. Frames, Concepts and Conceptual Fields. Lehrer, A. and E. Kittay (Eds.). 1992: 21-75.
- Barsalou, L.** 2003. Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes* 18(5-6): 513-562.
- Barsalou, L. and K. Wiemer-Hastings.** 2005. Situating Abstract Concepts. Pecher, D. and R. Zwaan (Eds.). 2005. *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thought*: 129-163. New York: Cambridge University Press.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Boguraev, B. and J. Pustejovsky.** 1990. Lexical Ambiguity and the Role of Knowledge Representation in Lexicon Design. Karlgren, H. (Ed.). 1990. *Proceedings of the 13th Conference on Computa-*

- tional Linguistics. Volume 2:* 36-41. <http://www.aclweb.org/anthology-new/C/C90/C90-2007.pdf>.
- Bolinger, D.** 1965. The Atomization of Meaning. *Language* 41: 555-573.
- Calzolari, N.** 1992. Acquiring and Representing Semantic Information in a Lexical Knowledge Base. *Lecture Notes in Computer Science* 627: 235-243.
- Calzolari, N., C. Fillmore, R. Grishman, N. Ide, A. Lenci, C. MacLeod and A. Zampolli.** 2005. Towards Best Practice for Multiword Expressions in Computational Lexicons. <http://gandalf.aksis.uib.no/lrec2002/pdf/259.pdf>.
- Carter, R.** 1997. Lexical Knowledge, Conceptual Knowledge, and Generativity. *Sémiotiques* 13: 41-64.
- Cimiano, P. and J. Wenderoth.** 2007. Automatic Acquisition of Ranked Qualia Structures from the Web. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*: 888-895.
- Crossley, S., T. Salsbury and D. McNamara.** 2009. Measuring L2 Lexical Growth Using Hypernymic Relations. *Language Learning* 59(2): 307-334.
- D'Andrade, R.** 1987. A Folk Model of the Mind. Holland, D. and N. Quinn (Eds.). 1987: 112-148.
- De Boni, M. and S. Manandhar.** 2002. Automated Discovery of Telic Relations for WordNet. *Proceedings of the 1st International WordNet Conference, India*.
- De Clerck, B., F. Verroens, D. Willems and T. Coleman.** To appear. The Syntactic Flexibility of (Recent) Verbs of Instrument of Communication: A Corpus-based Study. *Functions of Language*.
- Elman, J.** 2009. On the Meaning of Words and Dinosaur Bones: Lexical Knowledge without a Lexicon. *Cognitive Science* 33: 547-582.
- Faber, P. and R. Mairal Usón.** 1999. *Constructing a Lexicon of English Verbs*. Berlin: Mouton.
- Fellbaum, C.** 1990. English Verbs as a Semantic Net. *International Journal of Lexicography* 3(4): 278-301.
- Fillmore, C.** 2003. Double-decker Definitions: The Role of Frames in Meaning Explanations. *Sign Language Studies* 3(3): 263-295.
- Fillmore, C. and B. Atkins.** 1992. Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. Lehrer, A. and E. Kittay (Eds.). 1992: 75-102.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: Sun PReSS.
- Hanks, P.** 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205-215. (Also included in Fontenelle, T. (Ed.). 2008. *Practical Lexicography. A Reader*: 125-134. Oxford: Oxford University Press.)
- Holland, D. and N. Quinn.** 1987. *Cultural Models in Language and Thought*. Cambridge: University of Cambridge Press.
- Jiménez Briones, R.** 2007. The Lexical Representation of the English Verbs of Feeling within the Lexical Constructional Model. Paper read at the 40th Annual Meeting of the Societas Linguistica Europaea, University of Joensuu, Finland, August 29–September 1, 2007. http://www.lexicom.es/drupa/files/RocioJimenez_SLE2007.pdf
- Jiménez Briones, R. and B. De Alba.** 2008. Lexical Presentation within the Lexical Construction Model: An Analysis of Verbs of *happiness* and *happening*. *RESLAI* 21: 129-146.
- Keesing, R.** 1987. Models, "Folk" and "Cultural". Paradigms Regained. Holland, D. and N. Quinn (Eds.). 1987: 369-393.

- Keil, F.** 1979. *Semantic and Conceptual Development: An Ontological Perspective*. Cambridge, MA: Harvard University Press.
- Keil, F.** 1981. Constraints on Knowledge and Cognitive Development. *Psychological Review* 88: 197-227.
- Kövecses, Z.** 2006. *Language, Mind, and Culture. A Practical Introduction*. Oxford: Oxford University Press.
- Lehrer, A. and E. Kittay (Eds.)**. 1992. *Frames, Fields, and Contrasts. New Essays in Semantic and Lexical Organization*. Hillsdale, NJ: Lawrence Erlbaum.
- Li, P.** 2009. Lexical Organization and Competition in First and Second Languages: Computational and Neural Mechanisms. *Cognitive Science* 33: 629-664.
- Liu, M. and Y. Wu.** 2003. Beyond Frame Semantics: Insight from Mandarin Verbs of Communication. http://www.fl.nctu.edu.tw/~mliu/papers/Beyond%20Frame%20Semantics_Insight%20from%20Mandarin%20Verbs%20of%20Communication.pdf.
- Malle, B.** 2005. Folk Theory of Mind: Conceptual Foundations of Social Cognition. <http://cogprints.org/3315/>.
- McRae, K., V. De Sa and M. Seidenberg.** 1997. On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology. General* 126(2): 99-130.
- Miller, G.** 1990. Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography* 3(4): 245-264.
- Miller, G. and C. Fellbaum.** 1991. Semantic Networks of English. *Cognition* 41: 197-229.
- Moreno, A.** 2007. A Semantic Analysis of Universal and Idiosyncratic Features of Induced Motion Verbs: From the Perspective of Language Typology. *Journal of Universal Language* 8: 23-49.
- Morillas, J.** 1997. The Cultural Cognitive Model: A Programmatic Approach. *Cuadernos de Filología Inglesa* 612: 53-63.
- Murphy, G. and D. Medin.** 1985. The Role of Theories in Conceptual Coherence. *Psychological Review* 92(3): 289-316.
- Nation, P.** 2000. Learning Vocabulary in Lexical Sets: Dangers and Guidelines. *Tesol Journal* 9(2): 6-10.
- Pustejovsky, J.** 1990. The Generative Lexicon. *Computational Linguistics* 14(4). <http://acl.ldc.upenn.edu/J/J91/J91-4003.pdf>.
- Pustejovsky, J.** 1991. The Syntax of the Event Structure. *Cognition* 41: 47-81.
- Pustejovsky, J., C. Havasi, J. Littman, A. Rumshisky and M. Verhagen.** 2006. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. <http://www.cs.brandeis.edu/~arum/publications/lrec-bso.pdf>.
- Santos, A., S.E. Chaigneau, W.K. Simmons and L.W. Barsalou.** To appear. Property Generation Reflects Word Association and Situated Simulation. *Language and Cognition*.
- Shipley, T. and J. Zacks (Eds.)**. 2008. *Understanding Events: From Perception to Action*. Oxford: Oxford University Press.
- Vinson, D. and G. Vigliocco.** 2002. A Semantic Analysis of Noun-Verb Dissociation in Aphasia. *Journal of Neurolinguistics* 15: 317-351.
- Vinson, D., G. Vigliocco, S. Cappa and S. Siri.** 2003. The Breakdown of Semantic Knowledge: Insights from a Statistical Model of Meaning Representation. *Brain and Language* 86: 347-365.
- Wu, L.-L. and L. Barsalou.** 2009. Perceptual Simulation in Conceptual Combination: Evidence from Property Generation. *Acta Psychologica* 132: 173-189.

Reflections on the Academic Status of Lexicography

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa, and Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (st@asb.dk)*

Abstract: Two main camps have been formed with regard to the philosophical and academic status of lexicography: one considering lexicography an independent scientific discipline, and the other opposing such a scientific status. This article discusses some of the arguments from the second camp and argues that lexicography should be considered an independent scientific discipline. The argumentation is based on the fact that the subject field of lexicography is different from the subject fields of any other discipline, including linguistics. In this sense, the concept of a lexicographical work is broader than the more reduced concept of a dictionary. Lexicographical works, including dictionaries, are considered cultural artefacts and utility tools produced in order to meet punctual information needs detected in society. In this way, they have during the millenniums covered almost all spheres of human activity and knowledge. The theory and science of lexicography should not focus on the differences regarding the specific content of all these works, but on aspects that unite them and are common to all of them. In this regard, some of the core characteristics of lexicography as an independent discipline are discussed together with its complex relation to other disciplines. Lexicographical theory is understood as a systematic set of statements about its subject field. Finally, the article argues that the fact that this theory may seem too abstract and difficult to some working lexicographers does not in itself invalidate its independent scientific status, although a close relation between theory and practice is recommended.

Keywords: LEXICOGRAPHY, LEXICOGRAPHICAL THEORY, FUNCTION THEORY, ACADEMIC STATUS OF LEXICOGRAPHY, INDEPENDENT STATUS OF LEXICOGRAPHY, INTERDISCIPLINARY VOCATION OF LEXICOGRAPHY, DICTIONARIES, LEXICOGRAPHICAL WORKS, UTILITY TOOLS

Opsomming: **Gedagtes oor die akademiese status van die leksikografie.** Twee hoofkampe het ontwikkel met betrekking tot die filosofiese en akademiese status van die leksikografie: een wat die leksikografie beskou as 'n onafhanklike wetenskaplike dissipline, en die ander wat so 'n wetenskaplike status teenstaan. Hierdie artikel bespreek sommige van die argumente van die tweede kamp en redeneer dat die leksikografie as 'n onafhanklike wetenskaplike dissipline beskou behoort te word. Die redenasie is gebaseer op die feit dat die onderwerpsveld van die leksikografie verskillend is van onderwerpsvelde van enige ander dissipline, insluitende die linguistiek. Hiervolgens is die konsep van 'n leksikografiese werk breër as die meer beperkte konsep van 'n woordeboek. Leksikografiese werke, insluitende woordeboeke, word beskou as kulturele artefakte en nutsgereedskap geskep om aan die onmiddellike inligtingsbehoefes te voldoen wat in die gemeenskap vasgestel is. Op hierdie manier het hulle gedurende die millenniums byna alle sferes van menslike bedrywigheid en kennis gedek. Die teorie en wetenskap van die leksi-

kografie behoort nie te fokus op die verskille wat betref die spesifieke inhoud van al hierdie werke nie, maar op aspekte wat hulle verenig en algemeen aan hulle almal is. In hierdie verband word sommige van die kerneienskappe van die leksikografie as 'n onafhanklike dissipline bespreek saam met sy komplekse verwantskap aan ander dissiplines. Leksikografiese teorie word verstaan as 'n sistematiese reeks stellings oor sy onderwerpsveld. Ten slotte redeneer die artikel dat die feit dat hierdie teorie te abstrak en moeilik kan lyk vir sommige praktiserende leksikograwe, nie sy onafhanklike wetenskaplike status ongeldig maak nie, alhoewel 'n noue verwantskap tussen teorie en praktyk aanbeveel word.

Sleutelwoorde: LEKSIKOGRAFIE, LEKSIKOGRAFIESE TEORIE, FUNKSIETEOORIE, AKADEMIESE STATUS VAN DIE LEKSIKOGRAFIE, ONAFHANKLIKE STATUS VAN DIE LEKSIKOGRAFIE, INTERDISSIPLINÊRE TAAK VAN DIE LEKSIKOGRAFIE, WOORDEBOEKE, LEKSIKOGRAFIESE WERKE, NUTSGEREEDSKAP

1. Introduction

Among scholars dealing with lexicography, there is an old and deep-rooted dispute as to the philosophical and academic status of their discipline. Two main camps have been formed: one that considers lexicography an independent scientific discipline, and another that opposes such a status. The first camp has up till now found two major expressions, the lexicographical *function theory* developed at the Centre for Lexicography at the Aarhus Business School, Denmark, and the systematic *general theory of lexicography* presented by the German scholar Herbert Ernst Wiegand. The second camp has, as is natural for a discipline not considering itself to be independent, crystallized in a variety of positions which may nevertheless be ordered into three main groups with no sharp boundaries. In the best academic tradition, most of the renowned lexicographical journals, like *Lexicographica*, *LexicoNordica* and *Lexikos*, publish contributions by scholars belonging to both camps, thus promoting a hopefully rich and productive discussion for the benefit of the discipline.

In their introduction to a recent book, Atkins and Rundell (2008: 4) place themselves in one of the three groups belonging to the second camp:

This is not a book about 'theoretical lexicography' — for the very good reason that we do not believe that such a thing exists. But that is not to say that we pay no attention to theoretical issues. Far from it. There is an enormous body of linguistic theory which has the potential to help lexicographers to do their jobs more effectively and with greater confidence.

Thus, Atkins and Rundell reject the very existence of any lexicographical theory, although they do not exclude the need for lexicographers to be guided by a theory, in this case linguistic theory, which, as it will be recalled, embraces many competing schools. This position is somewhat surprising, especially because one of the two authors, Atkins (1992: 4-5), in a contribution reprinted in

Fontenelle (2008), does not deny the possibility of the concept of "theoretical lexicography":

Does theoretical lexicography exist? If this forum is to discuss its nature, we must at least grant it the benefit of the doubt ... I propose here to interpret the term theoretical lexicography in the very general sense of "a body of theory related to lexicography"... Although theoretical lexicography is certainly not synonymous with lexicology ..., nor with its daughter science lexical semantics ..., a large proportion of the decisions made by the lexicographer are linguistic decisions, and so we should consider particularly, but not exclusively, the contribution of linguistics to theoretical lexicography, and hence the role of the theoretical linguist in dictionary-making.

What is interesting here is the formulation "particularly, but not exclusively". This formulation seems to open the door for an understanding of "theoretical lexicography" as something more than just "theoretical linguistics", although this "something" is not clearly defined. We will later return to the claim that "a large proportion of the decisions made by the lexicographer are linguistic decisions", but here it is sufficient to note that the above quotation from Atkins and Rundell (2008) shows that the two authors do not deny the need for theoretical guidance, although they refer the corresponding assistance to the sphere of existing linguistic theory.

Another group of scholars, who are much closer to Atkins (1992) than to Atkins and Rundell (2008), regard lexicography as a sort of subdiscipline of linguistics and are, among others, represented by Ščerba (1940), who called for a "general theory of lexicography", though embedded in linguistics, and Reinhard Hartmann who, in several contributions, has argued for the need to develop a lexicographical theory, frequently called *metalexicography*. It is by no means surprising that Hartmann (2009: 90) criticizes the position taken by Atkins and Rundell (2008) on the relation between theory and practice as being full of simplifications and contradictions and that he is especially annoyed because the two authors "deprecate *metalexicography* as the province concerned merely with talking 'about dictionaries' ". In the same group can also be included scholars like Piotrowski (2009: 485) who advocates a "new theory on lexicography" strongly embedded in modern linguistics:

A new theory of lexicography would be one that would take seriously what we know about pragmatics and discourse, text structure, and would account for the contribution of particular textual elements to the dynamic meaning of a text, in short, the dynamics of meaning construction both in the receptive and the productive mode.

Somewhere in the midstream between the two previous groups are other lexicographers who consider themselves to be applied linguists. Meier (2003: 307), for instance, writes that "dictionary making, though often regarded as 'a special technique rather than a branch of linguistics', may be considered an instance of applied linguistics". In the same spirit can be mentioned Haensch et al. (1982), a

book that is simply entitled *Lexicography: From Theoretical Linguistics to Practical Lexicography*, Landau (2001) who also discusses lexicography from the point of view of linguistics referring to it as *The Art and Craft of Lexicography*, as well as Ten Hacken (2009: 399) who argues that linguistics, and especially Chomskyan linguistics, should be "taken as a background" in order to consider lexicography "a scientific activity".

A third group of authors denying the independent status of lexicography is represented by the lexicographer Urdang who, on more than one occasion, has rejected any possibility of being theoretically guided when involved in practical dictionary-making. The same position is shared by Wierzbicka (1985: 5) who claims that "lexicography has no theoretical foundation, and even the best lexicographers, when pressed, can never explain what they are doing, or why". The position taken by Urdang and Wierzbicka may be considered a sort of capitulation to the theoretical problems and challenges facing modern lexicography.

To summarize: It seems reasonable to distinguish between three main groups of lexicographers who deny their discipline's independent scientific status: one group that does not accept any theory at all; a second group that defends the development of a lexicographical theory, or metalexicography, embedded in linguistics; and a third group claiming that practical lexicography must be guided directly by linguistic theory. It is also clear that there are no sharp boundaries between the various positions which are frequently, as Hartmann rightly states, characterized by internal contradictions. The three groups comprise a large part of the lexicographers working with general dictionaries, lexicographers who frequently have their academic background in linguistics.

Whatever one thinks of their arguments, the very denial of any independent status to their discipline implies that lexicography as a whole *cannot* at the present moment be considered a totally independent discipline. But this does not mean that it *should* not be treated as such. As mentioned above, there is also another camp of lexicographers who do consider their discipline to be independent. One of the main expressions of this camp is the work of Wiegand (1989, 1998), who prefers the term *dictionary research* (German: *Wörterbuchforschung*) to the term *theoretical lexicography* when discussing the academic status of the discipline. However, in order to keep this article within acceptable limits, the following reflections will only deal with the other main expression, i.e. the lexicographical function theory.

2. Some basic considerations

The term *lexicography* is far from clear and unambiguous. One way to pinpoint its meaning is to go back in history and determine the origin of the term, which can be found in the Greek words *léksis* that means "speech, word", *leksikón* which is a book "dealing with words" and *gráphein* that means "to write". In the

light of this, lexicography would mean something like "the writing about words", i.e. a very broad definition which embraces any kind of text related to linguistics and dealing with language. However, at a certain moment in history, the terms *lexicography* and *lexicographer* became specifically related to the writing and writer of dictionaries. We do not know exactly when the two terms were first used in these senses, but undoubtedly it happened centuries ago as can be seen in the definition provided by Johnson (1755) in his *Dictionary of the English Language*:

Lexicographer A writer of dictionaries; a harmless drudge, that busies himself in tracing the original, and detailing the signification of words.

This definition of a lexicographer provided by Johnson is far from adequate. It goes without saying that lexicographers do "write" dictionaries but they also write other types of similar "books". In fact, only four years before the publication of Johnson's dictionary, the first volume of the French *Encyclopédie* was published with the complete title: *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* (*Encyclopedia, or a Systematic Dictionary of the Sciences, Arts, and Crafts*). It is noteworthy that the editors had decided to include both the word *encyclopedia* and the word *dictionary* in the very title of the book. The reason for this was not that they did not know the meaning of the two words because, among the total of 71 818 articles included in the 35 volumes, there is one about encyclopedias and another about dictionaries, written by each of the two main editors, d'Alembert (1754) and Diderot (1755), respectively.

Historically, in fact, lexicography does not only comprise dictionaries, but also encyclopedias, lexica, thesauri, glossaries and a number of other types of reference works. Recently, theoretical lexicography has even started to deal with how-tos, handbooks, manuals, etc. (Tarp 2007: 177). The present-day meaning of the term *lexicography* cannot be traced by means of etymology or determined through definitions given by lexicographers at different moments in history. The exact scientific meaning of the word *lexicography* is continuously changing and can only be determined on the basis of an up-to-date knowledge of the theoretical and practical development of this branch of human activity.

Both the French *Encyclopédie* and Johnson's *Dictionary of the English Language* may be considered old lexicographical works. But in a historical perspective their 250 years make them relatively young compared to the first-known dictionaries which, as it is generally accepted, were compiled more than 4 000 years ago. Since then, hundreds of thousands of different dictionaries and other lexicographical works have been published as a response to very different kinds of needs occurring in different cultures and periods. In this respect, Al-Kasimi (1977: 1) writes:

Dictionaries have developed not as theoretical instruments, but as practical tools. The major motives behind the rise of lexicography differ from one culture to another. Each culture fosters the development of dictionaries appropriate to its characteristic demands.

During the centuries, these practical tools have been imprinted on clay, hand-written on papyrus or paper, type-written, printed with different technologies or, more recently, made available on compact disks, the Internet and other electronic media. They have been written in hundreds of different languages, and even in dialects. They cover almost all spheres of human activity and knowledge. They have a huge variety of purposes. They transmit knowledge between generations and within each generation. They help people to communicate in their mother tongue and between and among language communities. They assist their users with advice and instructions in order to perform manual or mental actions. In short, dictionaries and other lexicographical works have played and still play an extremely important role in our cultural and social development. To reduce this rich cultural heritage to some sort of applied linguistics or, even worse, to something that cannot be described and treated theoretically is close to barbarism.

3. Independent subject field

A major criterion and necessary precondition for granting an independent status to lexicography is that its subject field is different from that belonging to any other discipline. Until now, the only discipline external to lexicography that has claimed property to lexicography is linguistics. The subject field of linguistics is language, something inherent in human beings without which they would not be human beings. Although the individual languages are historical and social products, modern man is genetically predisposed to develop language skills as a means of social communication, conscious thinking and knowledge storage. Contrary to this, the subject field of lexicography is constituted by dictionaries and other lexicographical works, i.e. purely cultural artefacts produced in order to satisfy various types of human needs as Al-Kasimi (1977) rightly points out in the above quotation.

The claim of Atkins (1992: 5) that "a large proportion of the decisions made by the lexicographer are linguistic decisions", does not correspond to the overall historical and present practice. This may, of course, be the case with some specific dictionaries compiled by lexicographers who import linguistic principles without letting them pass through any independent lexicographical "filter", but it cannot be generalised. Hence, the study of the "characteristic demands" of each culture giving rise to the development of dictionaries is hardly a linguistic discipline. Even more so because the satisfaction of many of these demands, or needs, does not require the co-operation of a linguist, but of other types of specialists. It is enough to take a glance at some of the 71 818 articles in the French *Encyclopédie*, which is also called a dictionary, in order to confirm this statement. In most cases, the authors of the individual articles have no specialised linguistic knowledge except for the skills that all native speakers have in their mother tongue. In this respect, it is useful to reread some

extracts from the article *Dictionnaire*, authored by d'Alembert (1754), who himself had no specialised linguistic training:

It is possible to distinguish between three types of *dictionary*: language dictionaries, historical dictionaries, and science and art dictionaries. This subdivision could be presented from a more general point of view as *dictionaries* of words, *dictionaries* of facts and *dictionaries* of things. Nevertheless, we retain the first subdivision because we believe that it is more convenient and even more precise. In fact, a language *dictionary*, which would seem to be nothing other than a word *dictionary*, when well done must often be a *dictionary* of things and is thus a very philosophical work. [...] A science *dictionary* can, and must, only be a *dictionary* of facts, whenever the causes are unknown to us, that is, almost always. [...] Finally, a historical *dictionary*, assembled by a philosopher, will often be a *dictionary* of things. Produced by an ordinary writer, by a compiler of facts and dates, it will be but a *dictionary* of words.

It is difficult to see the role of a linguist when it comes to producing what d'Alembert calls "historical dictionaries" and "science and art dictionaries" related to "facts" and "things". This especially becomes clear when we study some modern dictionaries dealing with specialised fields of human knowledge. In this respect, it is somehow surprising that the *International Journal of Lexicography* is one of the theoretical journals that almost exclusively publish articles strongly defending the subordination of lexicography to linguistics. The journal belongs to Oxford University Press which is also responsible for the publication of a large number of excellent specialised dictionaries where it is difficult to see any relation whatsoever to linguistic theory. One of these dictionaries is the *Concise Oxford Dictionary of Archaeology* compiled by Timothy Darvill, Doctor and Professor of Archaeology. In his preface, Darvill (2008: vii) writes about the content and the intended user group:

This dictionary of terms is intended as a guide to words likely to be commonly encountered in the archaeological literature or in the presentation of archaeological material through lectures, television, or video films. It is thus intended primarily for those who are essentially unfamiliar with the vocabulary and technical terms used by archaeologists, for example students starting out on a career in archaeology or seeking to expand their areas of special interest, journalists popularizing the scholarly outpourings of archaeology, and other professionals who have cause to dip into the archaeological literature.

Thus, this dictionary is, according to its author, conceived for interested laypeople (journalists), semi-experts (students) and professionals who seek assistance to cope with unfamiliar terms or to expand their areas of special interest. In terms of design and content, the articles included are similar to those that can be found in thousands of other specialised dictionaries:

Gallo-Belgic ware [Ar] Vessels imported from Gaul in the late 1st century BC and early 1st century AD, usually in black or silver-grey fabrics (**terra nigra*) or white fabric coated with red slip (**terra rubra*), or a

dense white or cream fabric like pipeclay. Close British imitations of these fabrics and forms are known, and further copying of the forms was widespread. The imported vessels often have the name of the potter stamped on the inner surface of the base, a practice imitated in Britain but usually with illegible markings.

It is difficult to see why the compilation of this dictionary should require special knowledge of linguistic theory. The author is not a linguist, but an expert in archaeology which is the topic of the dictionary. Of course, he is supposed to master general and archaeological English at a very high level and, in this respect, he can be considered a person with general linguistic knowledge. But it is not necessary to have studied linguistics and know anything about linguistic theory in order to conceive and produce the dictionary in question. On the contrary, absolutely necessary is specialised and up-to-date knowledge about archaeology (and this also implies knowledge about the discipline's specialised vocabulary because a scientific discipline cannot be learned without simultaneously learning its particular terminology). The need for this specialised, up-to-date knowledge can be seen not only in the above article, but also in the following observation by Darvill (2008: x):

Archaeological interpretations are constantly changing in the light of new evidence and ongoing research; in some spheres of interest such change can be quite rapid. Accordingly, the entries included here are point-in-time statements reflecting available knowledge and accepted understandings of what are undoubtedly complicated phenomena.

As mentioned in the previous section, it is a historical fact that thousands of dictionaries like the *Concise Oxford Dictionary of Archaeology* and the ones that d'Alembert calls "historical dictionaries" and "science and art dictionaries" related to "facts" and "things" have been produced during the last 4 000 years. To claim that the lexicographical decisions taken by the authors of these dictionaries are, to a large proportion, "linguistic decisions" is simply false. Nobody denies the important contributions made by linguists to many excellent dictionaries, but what is at stake here is the claim of linguistics for property of lexicography. The world has known quite a number of dictionaries and other lexicographical works for the production of which no specialised knowledge of linguistic theory was needed or used. This fact alone indicates that lexicography in general cannot be scientifically categorised as applied linguistics or a subdiscipline of linguistics, and it also makes it pointless to claim that "a new theory of lexicography" should be based on knowledge about pragmatics, discourse, text structure, etc. as claimed by Piotrowski (2009).

It is therefore necessary to approach the status of lexicography in another way. In this respect, two alternative approaches are possible: one viewing lexicography as being simultaneously the subdiscipline of a large number of disciplines, one of them linguistics; and another regarding it as a separate and completely independent discipline, i.e. a discipline with its own independent

subject field and core. The first approach leads to a schizophrenic understanding of lexicography, which can only explain the differences between the various lexicographical works, but not what unites them and is common to all of them. This schizophrenic approach should therefore be rejected. Consequently, if one does not want to join the capitulationist position of Urdang and Wierzbicka, according to which no theoretical explanation of and guidance to practical dictionary-making is possible, the only option left, justified by its specific subject field, is to treat lexicography as an independent scientific discipline characterised by a big interdisciplinary vocation. Lexicography is what is left when an abstraction is made from the specific content of the data taken from other disciplines and incorporated into its individual products.

4. Utility tools

The most fundamental aspects characterising all types of lexicographical works is that they are utility tools which implies that they have been produced in order to meet certain types of human needs detected in society. Dictionaries were originally invented with such needs in mind, and this fundamental relation between cultural and social needs on the one, and lexicographical works on the other hand, has not changed over the millenniums. The needs are not static, but vary from culture to culture, from epoch to epoch. And even within the same culture and epoch they cannot be viewed in an abstract way, since they are always closely related to concrete persons finding themselves in concrete situations. The same person may have different types of needs in different situations, while two persons may have different needs although they find themselves in the same type of situation. The study and typologisation of these needs in their close and dialectical relation to types of users and types of social situations are therefore the starting point for any general theory of lexicography considered as an independent branch of human scientific activity. This basic idea is what is expressed in the lexicographical function theory (cf. Bergeholtz and Tarp 2002, 2003, 2004, and Tarp 2008). It places lexicography as an independent scientific discipline within the broader framework of the social sciences.

It is a matter of course that not all human needs are lexicographically relevant. Material and physical needs cannot be met by lexicographical works, and neither can all types of psychological and mental needs. On the other hand, it is evident that the needs that can actually be satisfied lexicographically are of a great variety, a fact which is reflected in the corresponding variety of existing dictionaries and other lexicographical works. In this respect, what is common to lexicographical needs, and what separates them from all other types of needs, is that they are always needs for information which subsequently may be used for a large variety of purposes. However, information needs may be met not only by lexicographical works, but also by other types of texts such as books, text books, magazines, newspapers, etc. It is therefore necessary to fur-

ther clarify the concept of lexicographically relevant needs. To do so, it is useful to look at the distinction, discussed by Hausmann (1977: 144), between global and punctual needs (questions) and develop it further:

The question is now: What are the dictionaries for? For what purpose are they used? Which dictionaries are particularly well suited for what functions? The simplest and most frequent type of dictionary usage is undoubtedly punctual consultation to answer a very particular question ... The second type of question, which we can call systematic in contrast to punctual, is much less frequent. It is much less common for dictionaries to be used to answer these questions.

It cannot be denied that there are some rare cases of "dictionaries" designed to meet global needs. Hausmann calls them "primary learning dictionaries" (German: *primäre Lernwörterbücher*), i.e. dictionaries that are supposed to be read from one end to another. However, such dictionaries may be considered exceptions which confirm the thesis that no sharp boundaries exist within the subject field of social sciences in general. On the contrary, the overwhelming number of dictionaries and other lexicographical works has been designed with another view, i.e. to be used as consultation tools to meet punctual information needs. This fact brings lexicography close to modern information science.

It is obvious that the distinction between *punctual* and *global* should not be viewed in the light of *short* and *long*, but in the broader framework of *part* and *whole*. In this respect, there are lexicographical works with long entries, such as the French *Encyclopédie*, and the Chinese *Yongle Dadian* (Xie 1408), the world's biggest lexicographical work ever compiled, which, although never published, consisted of no less than 11 095 volumes partially transcribed character for character as exact copies of original texts written during the previous decades and structured according to a rhyming system for the characters, which made it possible to use the work for lexicographical consultation.

To summarize: Lexicography deals with utility tools designed to be consulted in order to satisfy punctual information needs related to specific types of users and specific types of social situations and varying from culture to culture, from epoch to epoch. In this sense, it seems very restrictive to reduce the concept of a lexicographical work to some general dictionaries to the production of which only linguists have contributed. The concept of a lexicographical work comprises all utility tools designed for this purpose and with these characteristics.

5. The core of lexicography

Contrary to general discourse, dictionaries do not contain information, but data from which their users may retrieve the information needed. This distinction between data and information, first introduced and discussed in detail by Wiegand (2000, 2002), is important because the very information retrieval is a mental and intellectual exercise whose result depends on the user's character-

istics in terms of culture, language proficiency level, and general or specialised knowledge. Two different types of users with the same type of information need may, for instance, require different types of data in order to retrieve the same information from lexicographical data. In this respect, it is necessary to point out that the concept of lexicographical data does not correspond to the concept of data resulting from empirical research, for example, corpus analysis. The lexicographical data are supposed to be selected, elaborated and presented according to lexicographical criteria with a view to making them understandable and, hence, useful to the specific type of user, i.e. to ensure that the information retrieval can actually be accomplished.

This understanding of lexicographical data implies that these data cannot be taken over uncritically from empirical research or directly from other disciplines, among them linguistics. If the data are not subjected to a meticulous selection, elaboration and presentation process according to lexicographical criteria, there is a high risk of reaching a low-quality lexicographical product, or at least a product with a number of problems and complications in terms of user-friendliness. This is the reason why it is recommendable to have a person with specialised lexicographical training responsible for the design of dictionary concepts, the writing of instructions, the preparation of working plans and the supervision of the practical lexicographical work.

Although it is also recommendable to engage experts from the respective subject fields in the practical work, they will not themselves have the necessary academic and scientific background allowing them to take overall lexicographical decisions and supervise the work with a view of ensuring a high-quality lexicographical work. This also indicates why it is problematic when Atkins (1992: 5) argues that "a large proportion of the decisions made by the lexicographer are linguistic decisions". Such a practice, even in dictionaries where specialised linguistic knowledge does have an important role to play, may lead to many unnecessary weaknesses and problems which could easily be avoided by involving a lexicographical expert in the project.

As mentioned in the previous section, what distinguishes lexicographical works from other kinds of texts which users may also use to satisfy their information needs is that they are consultation tools focusing on punctual needs. This implies that the core of lexicography does not only consist of the principles of data selection, elaboration and presentation, based on a typology of needs related to types of users and social situations. In order to fulfil their function as consultation tools, lexicographical works must also cater for a quick and easy access to the relevant data, in printed dictionaries by means of lexicographical structures and references, and in electronic ones by means of search engines and links. Hence, the principles of quick and easy data accessibility, equally related to specific types of user needs, is therefore part and parcel of the core principles integrating the independent theory of lexicography (cf. Bergeholtz 2009a, and Tarp 2009a).

6. The relation to other disciplines

Lexicography should be viewed not only as an independent scientific discipline in general, but also as a discipline particularly characterised by its big interdisciplinary vocation and its co-operative and integrating nature. As already mentioned, during the last 4 000 years lexicographical works have covered almost all spheres of human activity and knowledge. Historically, lexicography has emerged as an academic discipline through a dialectical interaction with other disciplines. In this complex process, it is natural that it has been influenced by these disciplines in terms of specific theories, concepts and methods.

However, the assimilation of these theories, concepts and methods borrowed from other disciplines has been highly problematic, giving rise to an endless series of uncertainties and disputes among lexicographers of the various schools and tendencies, for instance on the concept of collocation and its selection and presentation in dictionaries. In many cases, especially when linguists claiming property to lexicography have been involved, the concepts, theories and methods have been taken over uncritically and directly from these other disciplines, mainly from linguistics. This method used to assimilate concepts, theories and methods developed within other spheres of human activity is, as mentioned, highly problematic. And if lexicography is considered an independent scientific discipline, it is completely contrary to the philosophical and methodological principles guiding the relation and interaction between independent disciplines. For this reason, Tarp (2008: 12) calls for a critical attitude in the relation to and interaction with other disciplines:

This necessary interaction with other disciplines does not mean that lexicography can automatically take over the arsenal of concepts, theories and methods used by these other disciplines. The fact that the object of study is delimited, and the fact that lexicography has its own independent core distinguishing it from other disciplines, mean that all these concepts, theories and methods must be subjected to critical analysis with a view to determining what should be rejected, what can be used, and how the useful factors can be adjusted and adapted to suit the particular nature of lexicography.

This critical assimilation of concepts, theories and methods used within other disciplines is not only necessary in terms of lexicographical theory, but has a big impact on the quality of practical lexicographical works. For instance, Bergenholtz (2009b) shows that the linguistic concepts of collocation and part of speech are inadequate when designing high-quality dictionaries. Tarp (2009b) discusses why learner's lexicography cannot simply take over the concepts of beginners, and intermediate and advanced learners used within linguistics and language didactics. Likewise, Tarp (2009c) argues that the linguistic distinction between homonymy and polysemy is irrelevant in the majority of lexicographical works. And finally, Xue (2010) advocates a special lexicographical approach to the linguistic concepts of nominal countability and uncountability in production dictionaries for Chinese learners of English. These are all reasons

why even practical lexicographical decisions cannot be handed over to linguists without specialised lexicographical instruction or training.

7. Conclusion

As mentioned in the introduction to this article, Atkins and Rundell (2008: 4) outright declare that they "do not believe" that a lexicographical theory exists. This statement does not only represent an "anti-theoretical stance" as Hartmann (2009: 93) rightly states. It also contradicts the fact that various more or less elaborated lexicographical theories, and even competing theories, exist and have existed for decades, and that these theories are actually guiding a number of lexicographers in their practical work. If Atkins and Rundell had said instead that they disagree with these theories, find them incomprehensible, or consider them useless for their purposes, then it would have been understandable. Maybe the explication should be found in the following quotation from Atkins (1992: 29):

For me, there are three kinds of theoretical linguists, who may or may not consider themselves theoretical lexicographers. The work of the first group is too theoretical, too abstract, or too difficult, or shows too little conception of what practical lexicography is all about, for it to be of any immediate use in dictionary-making. I believe most working lexicographers share this opinion.

This point of view seems to ignore the complex relation between theory and practice, between a science and its practical applications. Nobody has ever claimed that a science or theory should necessarily be easy to understand and applicable for immediate use. As a rule, there are a number of mediations and intermediate levels between the most abstract parts of a science or theory and its practical applications. The majority of workers at a shipyard will probably find the mathematical principles and calculations included in the science of engineering too theoretical, too abstract, and too difficult for them to be of any immediate use in their daily work. But this does not mean that they reject the value of engineering and the technical instructions they receive from their foremen and instead start building advanced ships exclusively based on their own experience and practical training.

It may be true that the majority of "working lexicographers share the opinion" that some of the lexicographical theories are too abstract and difficult. They are perfectly allowed to do so, and it should not be considered a problem or obstacle to their work. In order to successfully perform their profession, they only need theory-based dictionary concepts, lexicographical instructions and detailed working plans which constitute the lowest mediating links between practical dictionary-making and the overall theory understood as a systematically structured set of statements about lexicographical works and their relation to social and cultural needs.

As Hartmann (2009: 90) rightly says, Atkins and Rundell (2008) "deprecate metalexigraphy as the province concerned merely with talking 'about dictionaries' ". Of course, it cannot be denied that a number of metalexigraphers, or theoretical lexicographers, has never been engaged directly in practical dictionary-making. But this fact in itself should not invalidate their theories, just as it does not invalidate the theories of the engineers if they do not work directly in production. However, a close dialectical relation between theory and practice is without any doubt both healthy and desirable for any branch of human activity. In this respect, it should be remembered that the lexicographical function theory, which considers lexicography an independent scientific discipline, has been developed at the Aarhus-based Centre for Lexicography which, in the last two or three decades, has been actively engaged in the production of more than fifty dictionaries, of which more than half have been directly designed and supervised by members of the Centre's staff. In this Centre, theory building goes along with the conception, planning and practical production of still new types of dictionaries. It is therefore necessary to use other arguments to deny the need for an independent theory of lexicography.

The discussion about the philosophical and scientific status of lexicography should not be viewed as a dispute between "good and bad" or as a controversy of merely academic interest. It should instead be understood and treated as a rich and rewarding discussion with large and immediate practical consequences. To approach lexicography as an independent discipline means to concentrate on the development of the aspects constituting the very core of lexicography and making it unique in relation to all other academic and scientific disciplines. It means to establish the correct relation between lexicography and society with a view to detect the real user needs occurring extra-lexicographically and to find the adequate lexicographical solutions. It means to focus on the development of the principles for selection, preparation and presentation of lexicographical data and the improvement of the various types of access routes to these data. And above all, it means to strengthen lexicography's interdisciplinary vocation and its relation and cooperation with all the other disciplines and sciences which, in one way or another, are relevant to lexicography and its practical products. This cooperation has existed for more than 4 000 years and there is no reason why it should not continue and be further strengthened in the future.

Literature

Al-Kasimi, Ali M. 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.

d'Alembert, Jean le Rond. 1754. Dictionnaire. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers*. Tome IV: 958-969. (= d'Alembert: Dictionnaire. Article de l'Encyclopédia. http://fr.wikisource.org/wiki/Article_Dictionnaire_de_l'Encyclop%C3%A9die)

Atkins, B.T. Sue. 1992-93. Theoretical Lexicography and its Relation to Dictionary-making. *Dictionaries. Journal of The Dictionary Society of North America* 14: 4-43.

- Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning. 2009a. Hurtig og sikker tilgang til informationer. *LexicoNordica* 16: 29-54.
- Bergenholtz, Henning. 2009b. Klassifikationen in der Linguistik und in der Lexikographie: Wortarten und Wortverbindungen. *Zeitschrift für Anglistik und Amerikanistik* 57(3): 275-288.
- Bergenholtz, Henning and Sven Tarp. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica* 18: 253-263.
- Bergenholtz, Henning and Sven Tarp. 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Bergenholtz, Henning and Sven Tarp. 2004. The Concept of "Dictionary Usage". Dollerup, Cai (Ed.). *Worlds of Words. A Tribute to Arne Zettersten*. *Nordic Journal of English Studies* 3(1): 23-36.
- Darvill, Timothy. 2008. *Concise Oxford Dictionary of Archaeology*. Second Edition. Oxford: Oxford University Press.
- Diderot, Denis. 1755. Encyclopédie. *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers*. Tome V: 635-648. (= Denis Diderot 1969: Oeuvres complètes. Édition chronologique. Tome 2. Paris: Club français du livre et société encyclopédique française.)
- Fontenelle, Thierry (Ed.). 2008. *Practical Lexicography: A Reader*. Oxford Linguistics. Oxford: Oxford University Press.
- Haensch, Günther, Lothar Wolf, Stefan Ettinger and Reinhold Werner. 1982. *La lexicografía. De la lingüística teórica a la lexicografía práctica*. Madrid: Gredos.
- Hartmann, Reinhard. 2009. [Review of] B.T. Sue Atkins and Michael Rundell. *The Oxford Guide to Practical Lexicography*. *International Journal of Lexicography* 22(1): 89-94.
- Hausmann, Franz Josef. 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Romanistische Arbeitshefte 19. Tübingen: Max Niemeyer.
- Johnson, Samuel. 1755. *Dictionary of the English Language*. London: J. & P. Knapton.
- Landau, Sidney I. 2001. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Meier, Hans H. 2003. Lexicography as Applied Linguistics. Hartmann, Reinhard (Ed.). 2003. *Lexicography. Critical Concepts III*: 307-318. London/New York: Routledge.
- Piotrowski, Tadeusz. 2009. [Review of] Sven Tarp. *Lexicography in the Borderland Between Knowledge and Non-knowledge*. General Lexicographical Theory with Particular Focus on Learner's Lexicography. *International Journal of Lexicography* 22(4): 480-486.
- Ščerba, Lev V. 1940. Towards a General Theory of Lexicography. *International Journal of Lexicography* 8(4): 315-350, 1995.
- Tarp, Sven. 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tarp, Sven. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. *Lexicographica Series Maior* 134. Tübingen: Niemeyer.
- Tarp, Sven. 2009a: Reflections on Data Access in Lexicographic Works. Nielsen, Sandro and Sven Tarp (Eds.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 43-62. Terminology and Lexicography Research and Practice 12. Amsterdam/Philadelphia: John Benjamins.
- Tarp, Sven. 2009b. The Foundations of a Theory of Learners' Dictionaries. *Lexicographica. International Annual for Lexicography* 25: 155-158.

- Tarp, Sven.** 2009c. Homonymy and Polysemy in a Lexicographic Perspective. Bergenholtz, Henning and Thomas Herbst (Eds.). 2009. *Classifications in Linguistics and Lexicography. Special Issue of Zeitschrift für Anglistik und Amerikanistik* 57(3): 289-306.
- Ten Hacken, Puis.** 2009. What is a Dictionary? A View from Chomskyan Linguistics. *International Journal of Lexicography* 22(4): 399-421.
- Wiegand, Herbert Ernst.** 1989. Der gegenwärtige Status des Lexikographie und ihr Verhältnis zu anderen Disziplinen. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*: 246-280. Handbücher zur Sprach- und Kommunikationswissenschaft, Band 5.1. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband.* Berlin/New York: De Gruyter.
- Wiegand, Herbert Ernst.** 2000. Wissen, Wissenrepräsentationen und Printwörterbücher. Heid, Ulrich, Stefan Evert, Egbert Lehmann and Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 15-38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Wiegand, Herbert Ernst.** 2002. Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen. Ezawa, Kennosuke, Wilfried Kürschner, Karl H. Rensch and Manfred Ringmacher (Eds.). 2002. *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums Berlin 1998*: 265-281. Tübingen: Gunter Narr Verlag.
- Wierzbicka, Anna.** 1985. *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma.
- Xie, Jin (Ed.).** 1408. *Yongle Dadian 1–11, 095.* China.
- Xue, Mei.** 2010. Countable or Uncountable? That is the Question — Lexicographic Solutions to Nominal Countability in Learner's Dictionaries for Production Purposes. *Lexikos* 20: 540-558.

Skoolwoordeboeke vir huistaalleerders van Afrikaans

Sven Tarp, *Departement Afrikaans en Nederlands, Universiteit van Stellenbosch, Stellenbosch, Suid-Afrika en Sentrum vir Leksikografie, Aarhus Sakeskool, Universiteit van Aarhus, Aarhus, Denemarke (st@asb.dk)*

en

Rufus H. Gouws, *Departement Afrikaans en Nederlands, Universiteit van Stellenbosch, Stellenbosch, Suid-Afrika (rhg@sun.ac.za)*

Opsomming: Skoolwoordeboeke vorm 'n uitgebreide tipologiese kategorie binne die versameling woordeboeksoorte. Die uiteenlopende verskeidenheid teikengebruikers van skoolwoordeboeke vereis egter 'n duideliker tipologiese omskrywing en 'n fyner onderverdeling in subkategorieë. Hierdie artikel fokus eerstens op sommige van die verskillende interpretasies van die term *skoolwoordeboek* en die onduidelikheid wat met die gebruik van hierdie term saamhang. 'n Onpresiese gebruik van die term lei daartoe dat voornemende gebruikers van skoolwoordeboeke nie weet wat om in 'n betrokke woordeboek te verwag en wie die beplande teikengebruikers daarvan is nie. Die leksikografiese funksies van skoolwoordeboeke word bespreek met die klem op Afrikaanse skoolwoordeboeke vir huistaalleerders. 'n Oorsig en kontekstualisering van hierdie woordeboeke word gevolg deur 'n bespreking van verskeie woordeboeke vir laerskool- en hoërskoolleerders. Aandag word gegee aan aspekte van die funksies, struktuur en inhoud asook aan positiewe en negatiewe kenmerke van die onderskeie woordeboeke. Dit blyk dat die ontwikkeling in die teoretiese leksikografie 'n positiewe invloed op die samestelling van skoolwoordeboeke gehad het. Uiteindelik word 'n aantal aanbevelings gemaak vir die verbetering van toekomstige skoolwoordeboeke. Die belangrikheid van 'n ondubbelsinnige aanduiding van onder meer die teikengebruikers en die gebruikerssituasie word benadruk. Die opname, aanbieding en bewerking van data moet met inagneming van die gebruikersprofiel geskied.

Sleutelwoorde: ADDISIONELE TAAL, GEBRUIKERSITUASIE, GRONDSLAGFASE, HOËRSKOOLLEERDERS, HUISTAAL, KOGNITIEWE FUNKSIE, LAERSKOOLLEERDERS, LEKSIKOGRAFIESE FUNKSIES, SKOOLWOORDEBOEK, TEIKENGEBRUIKER, TEKSPRODUKSIE, TEKSRESEPSIE, WOORDEBOEKTIPOLOGIE

Abstract: **School Dictionaries for Home Language Learners of Afrikaans.**

School dictionaries constitute an extended typological category within the collection of dictionary types. The divergent variety of target users of school dictionaries however demands a clearer typological description and a more precise division in subcategories. This article focuses firstly on some of the different interpretations of the term *school dictionary* and the obscurity related to the use of this term. An imprecise use of the term leads to a situation where potential users of school dictionaries do not know what to expect in a specific dictionary and who the envisaged target users of the dictionary are. The lexicographic functions of school dictionaries are discussed with the emphasis on Afrikaans school dictionaries for home language learners. An overview and contextualisation of these dictionaries are followed by a discussion of different dictionaries for primary and

secondary school learners. Attention is drawn to aspects of the functions, structure and contents as well as positive and negative features of the different dictionaries. It becomes evident that the development in theoretical lexicography has had a positive influence on the compilation of school dictionaries. In conclusion a number of suggestions are made for the improvement of future school dictionaries. The importance of an unambiguous indication of among others the target users and the user situation is stressed. The inclusion, presentation and treatment of data must be done in accordance with the user profile.

Keywords: ADDITIONAL LANGUAGE, COGNITIVE FUNCTION, DICTIONARY TYPOLOGY, FOUNDATION PHASE, SECONDARY SCHOOL LEARNERS, HOME LANGUAGE, LEXICOGRAPHIC FUNCTION, PRIMARY SCHOOL LEARNERS, SCHOOL DICTIONARY, TARGET USER, TEXT PRODUCTION, TEXT RECEPTION, USER SITUATION

1. Inleiding

In sy informatiewe boek oor pedagogiese leksikografie wys Welker (2008) daarop dat die grootste deel van die literatuur oor hierdie onderwerp gerig is op woordeboeke vir leerders van 'n niemoedertaal (vgl. Tarp 2010). Eintlik sou 'n mens die omgekeerde situasie te wagte wees, aangesien daar baie meer potensiele gebruikers van woordeboeke vir moedertaalsprekers is as woordeboeke vir leerders van vreemde tale. Hierdie verrassende feit sluit dit egter nie uit nie dat die algemene terminologiese verwarring en onsekerheid wat kenmerkend is van leksikografiese literatuur oor die algemeen, ook tot die sfeer van pedagogiese leksikografie vir moedertaalsprekers deurgedring het. Een van die gevolge van hierdie ontwikkeling is die omstredenheid van die begrip *skoolwoordeboek*. Sommige navorsers soos Hartmann en James (1998) en Welker (2008) beskou 'n skoolwoordeboek as 'n woordeboek wat ontwerp is om te help met die aanleer van 'n eerste of moedertaal, d.w.s. 'n huistaal. Hernandez (1989) en Lagane (1989) wat onderskeidelik net oor Spaanse en Franse skoolwoordeboeke praat, deel oënskynlik hierdie standpunt. Lombard (1990) rig sy bespreking ook volledig op eentalige skoolwoordeboeke terwyl Nyangone Assam (2006) en Mabika Mbokou (2006) fokus op skoolwoordeboeke in Gaboen met 'n hibriediese struktuur: eentalig met 'n tweetalige dimensie of tweetalig maar steeds gerig op huistaalsprekers, selfs van albei die behandelde tale. Daarteenoor bied ander navorsers 'n breër omskrywing in ooreenstemming met die praktyk in verskillende lande met onder meer tweetalige skoolwoordeboeke wat duidelik nie ontwerp is as 'n hulpmiddel in die aanleer van 'n eerste taal nie, asook gespesialiseerde skoolwoordeboeke wat hulp moet verleen met die aanleer van 'n spesifieke vakgebied.

Om ook aan hierdie praktyk erkenning te gee, moet die gebruik van die term *skoolwoordeboek* uitgebrei word om te verwys na enige woordeboek wat saamgestel is om in die skool gebruik te word met die oog daarop om ondersteuning te bied aan die aanleer van 'n huistaal, 'n addisionele taal of 'n vakgebied. In hierdie opsig is *skoolwoordeboek* dan 'n generiese term wat verskillende

tipes woordeboeke insluit. Elkeen van hierdie tipes moet nader omskryf word om 'n presiese terminologie te verkry; ten minste vir gebruik in die teoretiese literatuur. Die begrip van 'n skoolwoordeboek sluit egter nie alle woordeboeke in wat op skool gebruik word nie maar slegs daardie woordeboeke wat ontwerp is vir daardie doel. In hierdie verband moet 'n verdere onderskeid gemaak word tussen skoolwoordeboeke wat verkorte of gereduseerde weergawes van omvattender algemene woordeboeke is en die *werklike* skoolwoordeboeke wat van meet af aan as sodanig beplan is en daarom aangepas is by die skoolleerders se taalkundige, verstandelike, kulturele en ensiklopediese ontwikkeling.

Wanneer ons skoolwoordeboeke omskryf as woordeboeke wat ontwerp is om op skool gebruik te word, verwys ons na *alle* woordeboeke wat beplan is om skoolleerders te help vanaf die begin van hulle skoolloopbaan op die ouderdom van 5, 6 of 7 totdat hulle skool verlaat op die ouderdom van 16, 17 of 18, na gelang van die betrokke land se onderwysstelsel. Hier word in ag geneem dat daar verwarring voorkom in die literatuur van die teoretiese leksikografie. Ter illustrasie: in sy artikel oor kinderwoordeboeke dui Hausmann (1990) aan dat daar geen duidelike skeidslyn tussen kinderwoordeboeke en skoolwoordeboeke bestaan nie en dat die titels dikwels nie die regte inhoud weerspieël nie. Vergelykbare kommentaar is te vinde in Hartmann en James (1998: 20) wat onder die lemma *children's dictionary* die volgende sê:

The transition between the dictionary for younger children and the school dictionary is fluid,

en in Bergenholtz et al. (1997: 214) wat onder die lemma *pedagogical dictionary* skryf:

Dit kan moeilik wees om 'n grens te stel tussen die mees elementêre skoolwoordeboeke (beginnerswoordeboeke) en kinderwoordeboeke.

Alhoewel Hausmann en Bergenholtz et al. die gebruikers van kinderwoordeboeke as tot en met die ouderdom van 10 beskou (terwyl ander navorsers dié ouderdom op 14 stel) definieer hulle, asook Hartmann en James, hierdie woordeboeke veral fenomenologies, dit is volgens die data wat in hulle ingesluit word. Indien die funksies van die onderskeie woordeboeke egter as vertrekpunt dien, word dit duidelik dat die meeste sogenaamde kinderwoordeboeke in feite beplan is om deur skoolkinders in die eerste grade van die laerskool gebruik te word en dat hulle spesifieke kenmerke die logiese resultaat van die skrywers se pogings is om dié woordeboeke aan te pas by die verstandelike en taalkundige ontwikkeling van hulle gebruikers. In hierdie verband moet die dikwels arbitrêre titels wat deur die uitgewers voorgestel word, nie beskou word as 'n relevante kriterium waarop 'n tipologie van woordeboeke gebaseer kan word nie; hulle werklike funksies is wel so 'n kriterium. Ons klassifiseer daarom die meeste sogenaamde kinderwoordeboeke as skoolwoordeboeke in-

dien dit op een of ander manier duidelik is dat die hoofgebruikersgroep van so 'n woordeboek laerskoolleerders is.

Die begrip *skoolwoordeboek* verwys nie slegs na afsonderlik gedrukte woordeboeke nie maar ook na woordeboeke, dikwels glossaria genoem, wat geïntegreer is as agtertekste in verskeie handboeke. Alhoewel hierdie tipe woordeboek feitlik volledig geïgnoreer word in die teoretiese literatuur, miskien met die uitsondering van De Sousa (1995: 143) wat woordeboeke wat terme definieer wat in sekere skoolhandboeke gebruik word, beskou as een van die twee basiese betekenisse van die term *skoolwoordeboek*, verdien hierdie glossaria tog baie meer teoretiese aandag omdat hulle belangrike leksikografiese werke is wat daaglik deur talle skoolleerders gebruik word. In hierdie artikel word daar voortaan net gewerk met afsonderlik gedrukte skoolwoordeboeke vir huistaalleerders van Afrikaans. 'n Bespreking en analise van handboekgeïntegreerde glossaria moet in 'n latere artikel aan die orde kom.

2. Funksies van skoolwoordeboeke vir huistaalsprekers

Skoolwoordeboeke vir huistaalleerders kan baie funksies hê. Dit geld met betrekking tot sowel die bepaalde gebruikersgroep wat vasgestel word deur die ouderdom, graad en die ooreenstemmende intellektuele, taalkundige, kulturele en ensiklopediese ontwikkeling van skoolleerders as die verskillende tipes leersituasies waar die kinders die woordeboek benodig of raadpleeg. Daar is eerstens twee kommunikatiewe situasies, dit is resepsie en produksie van geskrewe en gesproke tekste, waarin skoolwoordeboeke hulle gebruikers kan ondersteun. Tweedens is daar twee fundamentele kognitiewe situasies, dit is die leer van woordeskat en die leer van grammatika, waar kinders, dikwels saam met die opvoeder, die materiaal wat in woordeboeke voorsien word, gebruik om hierdie twee basiese komponente van taal te bestudeer. Aangesien 'n mens nie 'n woord kan leer sonder om te weet waarna dit verwys nie, val woordeskatleer dikwels, veral by jong kinders, saam met die leer van ensiklopediese en kulturele sake. In hierdie verband kan skoolwoordeboeke kinders ook bystaan in 'n derde kognitiewe situasie, dit is om meer van die wêreld te leer en wêreldkennis te verwerf as 'n basis vir woordeskatleer. Hier vind 'n aktiewe wisselwerking tussen woordeskat- en wêreldkennis plaas. Woordeboeke wat spesifiek beplan is om te help met woordeskatleer bevat gewoonlik tematiese afdelings, dikwels met prentillustrasies, soms selfs as "lemmas", terwyl woordeboeke wat met grammatikaleer help spesiale afdelings bevat, in gedrukte woordeboeke meestal in die voorteksteafdeling, waar verbuiging, woordvorming, leestekengebruik en ander grammatiese aspekte op 'n sistematiese manier behandel word. In hierdie gevalle word die skoolwoordeboeke nie primêr as raadplegingsinstrumente gebruik nie maar eerder as minihandboeke wat, in plaas van "gewone" handboeke, afdeling vir afdeling bestudeer kan word.

Afgesien van die gemelde twee kommunikatiewe en drie kognitiewe situasies wat almal van regstreekse belang vir die leer van taal is, kan skoolwoordeboeke in sommige lande ook hulp bied in 'n ander tipe kognitiewe situasie waar kinders iets van hulle taal moet weet. Dit is veral die geval wanneer die nasionale kurrikulum, byvoorbeeld in Suid-Afrika, vereis dat skoolleerders in spesifieke grade iets oor die herkoms en geskiedenis van hulle taal en sy woorde moet leer. Die gevolglike etimologiese funksie van sommige Afrikaanse skoolwoordeboeke hou dus nie verband met die leer van die taal nie maar net met die verwerwing van geleerde kennis oor die taal.

Ten slotte kan sommige skoolwoordeboeke, veral die sogenaamde kinderwoordeboeke, 'n bykomende onderliggende funksie hê, naamlik om skoolleerders te help om die gewoonte om woordeboeke te gebruik aan te leer en om woordeboekgebruiksvaardighede te ontwikkel (in Tarp 2008 word meer oor operasionele situasies gesê). In sy Spaanse leksikografiewoordeboek oordryf De Sousa (1995: 158) miskien hierdie funksie wanneer hy 'n kinderwoordeboek beskryf uitsluitlik as 'n woordeboek wat spesiaal beplan is om kinders in te lei in die gebruik van hierdie tipe werk.

Opsommend kan gesê word dat skoolwoordeboeke vir huistaalleerders die volgende sewe kommunikatiewe, kognitiewe en operasionele funksies behoort te hê waarvan slegs vyf regstreeks verband hou met die leer van die huistaal/ eerste taal:

Kommunikatiewe funksies regstreeks ter sake vir taalleer

- (1) Om skoolleerders te help met teksresepsie (skriftelik of mondeling)
- (2) Om skoolleerders te help met teksproduksie (skriftelik of mondeling)

Kognitiewe funksies regstreeks ter sake vir taalleer

- (3) Om skoolleerders te help met die leer van grammatika
- (4) Om skoolleerders te help met woordeskatleer
- (5) Om skoolleerders te help om van die wêreld te leer

Kognitiewe funksies nie regstreeks ter sake vir taalleer nie

- (6) Om skoolleerders te help om etimologie te leer

Operasionele funksie nie regstreeks ter sake vir taalleer nie

- (7) Om skoolleerders te help met die ontwikkeling van woordeboekvaardighede

Dit is vanselfsprekend dat verskeie van hierdie funksies beperk is tot skoolleerders van 'n sekere ouderdom en dat die individuele skoolwoordeboek, selfs al is dit van hoogstaande gehalte, nie daardie funksies hoef te vertoon wat nie ter sake is vir die ouderdom en graad van die beplande gebruikersgroep nie.

3. Afrikaanse skoolwoordeboeke: oorsig en kontekstualisering

Een van die tipiese probleme in baie omgewings waar woordeboeke gebruik moet word, is dat gebruikers nie genoegsaam bewus is van die verskeidenheid woordeboeksoorte wat bestaan nie. Daar word te dikwels gewerk met die idee van "die woordeboek" in plaas van "'n woordeboek" en dit lei daartoe dat een woordeboek gesien word as 'n bron waarin oplossings vir alle leksikografies verwante probleme gesoek moet word. Die skep van 'n woordeboekkultuur bring groter kennis en bewustheid van tipologiese verskeidenheid mee. 'n Woordeboekkultuur wat volgens Hausmann (1989) daartoe lei dat die gemeenskap by die leksikografie aanpas, is nie net ter sake vir bestaande en voornemende woordeboekgebruikers nie. Ook leksikografe toon dikwels te weinig tekens van die invloed van 'n woordeboekkultuur. Dit blyk veral daaruit dat daar in die leksikografie met sy breë tipologiese verskeidenheid dikwels onduidelikheid bestaan oor die presiese aard en omvang van 'n bepaalde tipologiese kategorie en oor hoe woordeboeke wat tot so 'n kategorie behoort, inas by die groter woordeboekversameling van 'n bepaalde taal. Hier gaan dit om sowel die teikengebruiker as die leksikograaf wat met dié onduidelikheid sukkel.

In die gesprek oor woordeboektipologie (vgl. onder meer Malkiel 1967, Zgusta 1971, Geeraerts 1984, Kromann et al. 1984, Gouws 1989, 2007) word dit duidelik gestel dat daar 'n verskeidenheid woordeboektypes is en dat elke tipe 'n bepaalde opdrag het, op 'n spesifieke tipe inligtingsoordrag gerig is en die data op 'n bepaalde manier moet verpak sodat die geïdentifiseerde teikengebruiker die inligting ten beste kan ontsluit. Dit word ook aanvaar dat woordeboeksoorte nie afgeslote kategorieë is nie maar dat daar oorvleueling tussen verskillende subtypes is. Die onderskeid is nie absoluut nie. In onlangse navorsing oor woordeboektipologie (vgl. Gouws 2007) word dit beklemtoon dat daar afgewyk moet word van 'n te rigiede tipologiese klassifikasie ten gunste van 'n soepeler indeling van woordeboeke volgens gebruikersbehoefes. So 'n pleidooi vir 'n soepeler indeling impliseer nie 'n minder duidelike aanduiding van die woordeboeksoort nie maar eerder 'n buigsamer benadering waar oorvleueling erken word en waar daar met andersoortige bepalende kenmerke eerder as suiwer linguïstiese kriteria gewerk word. So 'n benadering moet egter waak teen die gebruik van te vae of te lomp tipologiese klassifikasies wat aan 'n voornemende gebruiker te min leiding gee oor wat hy/sy in die betrokke woordeboek kan verwag en wat aan die leksikograaf te veel geleentheid gee om op 'n niespesifieke manier met die opname en bewerking van woorde om te gaan.

Die kategorie skoolwoordeboeke is juis 'n voorbeeld van so 'n tipologiese kategorie wat dikwels deur vaagheid en lompheid gekenmerk word. Aangesien skoolwoordeboeke met 'n duidelik geïdentifiseerde maar onderling uiteenlopende teikengebruikersgroep werk, is dit belangrik om 'n meer presiese omskrywing van die gebruiker te bied eerder as om bloot te sê dat 'n bepaalde

naslaanwerk 'n skoolwoordeboek is. Gegee die feit dat 'n leerder in Suid-Afrika se skoolloopbaan oor minstens 12 jaar strek en dat die taalbehoefte van leerders uiteenlopend is, behoort dit duidelik te wees dat die tipologiese kategorie skoolwoordeboeke geensins homogeen is nie. 'n Belangrike onderskeid binne hierdie kategorie is dié tussen woordeboeke wat die algemene taal en dié wat gespesialiseerde taal as onderwerp het. Binne albei hierdie kategorieë kan daar ook 'n onderskeid gemaak word tussen eentalige en twee- of meertalige woordeboeke en daarnaas ook eentalige woordeboeke wat 'n tweetalige dimensie het. Laasgenoemde tipe kom veral voor wanneer 'n woordeboek wat primêr as eentalige woordeboek saamgestel is, ook bykomende inskrywings, tipieserwys vertaalekwivalente, in 'n tweede taal bevat. Binne elk van die genoemde kategorieë is onderverdelings aantoonbaar en dié verskillende subtypes kan veral gemotiveer word op grond van teikengebruikers en gebruikersbehoefte. 'n Primêre onderskeid is tussen woordeboeke vir huistaalsprekers en vir tweede- of addisionele taalsprekers. Daarbenewens is die onderskeid tussen woordeboeke vir verskillende ouderdomme of grade in die skool ook van wesenlike belang. 'n Bepalende tipologiese aspek wat in al hierdie subtypes 'n rol behoort te speel, is die leksikografiese funksie van die betrokke woordeboek. Die verskillende tipes skoolwoordeboeke sal veral van mekaar verskil ten opsigte van 'n kognitiewe al dan 'n kommunikatiewe funksie en binne laasgenoemde groep sal tussen teksresepsie en teksproduksie onderskei word (vgl. Tarp 2000). Daarnaas moet 'n onderskeid tussen mono-, bi- en polifunksionele woordeboeke ook die tipologiese subverdeling binne die kategorie skoolwoordeboeke beïnvloed.

Woordeboekgebruikers moet besef dat een skoolwoordeboek nie die leksikografiese behoeftes van alle leerders kan bevredig nie en behoort daarom met groter omsigtigheid 'n woordeboekkeuse uit te oefen. Eweneens moet leksikograwe wat skoolwoordeboeke skryf, 'n subtipologiese sensitiwiteit aan die dag lê en 'n ondubbelsinnige aanduiding gee ten opsigte van die spesifieke gerigtheid van die woordeboek asook wie die geïdentifiseerde teikengebruikersgroep is. Dit mag nie aan die gebruiker oorgelaat word om self te moet besluit wie die teikengebruiker is nie. Ongelukkig laat te veel leksikograwe na om hierdie leiding in die titel, subtitel of elders in hulle woordeboek weer te gee. Te dikwels word dit nie eers eksplisiet gestel dat dit om 'n skoolwoordeboek gaan nie.

Waar woordeboeke wat vir skole saamgestel word, wel in byvoorbeeld die titel, subtitel of opdragomskrywing eksplisiet daarop dui dat die betrokke woordeboek vir gebruik in skole saamgestel is, word so 'n aanduiding te ongereeld vergesel van 'n identifisering van die spesifieke ouderdomsgroep of skoolgraad waarop die woordeboek gerig is. Binne die Afrikaanse en Suid-Afrikaanse konteks word daar te dikwels nog te maklik met die onskerp klassifikasie "skoolwoordeboek" gewerk. Hierdie klassifikasie misken die feit dat geen enkele skoolwoordeboek geskik vir alle grade en ouderdomsgroepe kan

wees nie. Veel groter sensitiwiteit moet aan die dag gelê word om te verseker dat die beplanning en samestelling van 'n woordeboek wat binne die breë kategorie van skoolwoordeboeke val, op die spesifieke behoeftes en naslaanvaardighede van 'n veel duideliker omskrewe teikengebruikersgroep binne die versameling leerders gerig is. Hier is 'n blote verdeling tussen gebruik in hoër- en gebruik in laerskole ook nie bevredigend nie aangesien daar byvoorbeeld in die laerskool 'n te groot verskil in die behoeftes en vaardighede van grondslagfaseleerders en leerders in die senior intermediêre fases bestaan. 'n Vraag kan met reg ook gevra word of 'n enkele woordeboek vir gebruik in hoërskole regverdigbaar is aangesien die graad 8-leerder en die graad 12-leerder uiteenlopende behoeftes het. Ook hier moet die moontlikheid ondersoek word om tipologies te onderskei tussen 'n woordeboek vir byvoorbeeld graad 8–10 en 'n ander vir graad 11–12.

In die beplanning van skoolwoordeboeke, veral in die Suid-Afrikaanse omgewing, moet daar ook deeglik rekening gehou word met die feit dat daar nie altyd 'n voorspelbare een-tot-een-afparing bestaan tussen die skoolgraad en die huistaalsprekersvlak of die skoolgraad en die onderrigtaalvlak van 'n leerder nie. Tipieserwys is 'n leerder van 11 miskien in graad 5 maar 'n woordeboek wat gebruik moet word om hom/haar met teksproduksie en/of teksrepsie te help, moet miskien eerder op graad 4-vlak gerig wees. Die werklike lees- en verstaanvermoë van 'n leerder is dikwels gelykstaande aan wat tot 'n graad of wat laer behoort. Wat vir die huistaalspreker van 'n bepaalde skoolgraad ter sake is in 'n gegewe woordeboek is nie onproblematies in 'n meertalige omgewing waar al die tale nie dieselfde beregtiging as onderrigtaal het nie. Daar moet naamlik onderskei word tussen woordeboeke vir onderrigtaal en woordeboeke vir huistaalonderrig aangesien die onderrigtaal nie noodwendig die huistaal van 'n betrokke leerder is nie. 'n Leerder kan dus in byvoorbeeld graad 6 wees maar die onderrigtaal is sy/haar tweede of derde taal wat egter as eerste taal deur die leerder gevolg moet word. Alhoewel so 'n leerder dan in graad 6 is, is die vlak van sy/haar kennis van die taal van onderrig wat ook in die betrokke skool as sy/haar eerste taal aangebied word, eerder gelykstaande aan dit wat minstens een vlak laer aangebied word. 'n Leerder in graad 8 sal dan byvoorbeeld veel meer baat by 'n woordeboek wat op graad 7, eerder as een wat op graad 8 gerig is. Dit word selfs vir taalgebruikers wie se huistaal hulle eerste taal en onderrigtaal is, bevestig deur bevindinge van Prinsloo en De Schryver (Op koms) dat gebruikers dikwels gemakliker is met 'n woordeboek wat op 'n vlak laer, eerder as een wat op hulle eie vlak gerig is.

In die Suid-Afrikaanse woordeboekemark sluit die kategorie skoolwoordeboeke ook 'n verskeidenheid tweetalige woordeboeke in. In die geval van byvoorbeeld 'n woordeboek met Afrikaans en Engels as behandelde taalpaar is dit miskien dikwels so dat die woordeboek (ook) op die Afrikaanssprekende leerder se behoeftes gerig is maar dan veral die behoeftes ten opsigte van Engels, dit is die tweede taal. In hierdie verband stel die *Oxford Afrikaans-*

Engels/English–Afrikaans Skoolwoordeboek/School Dictionary dit heel duidelik op sy agterblad dat die woordeboek ontwerp is "om leerders doeltreffend en bedrewe in hulle addisionele taal te laat skryf en praat". So 'n woordeboek is dus nie gerig op die behoeftes van die huistaalspreker van Afrikaans ten opsigte van Afrikaans nie. Gevolglik val dit buite hierdie bespreking. 'n Tweetalige skoolwoordeboek wat vir baie lank 'n dominante rol in die skoolmark gehad het, is die *Tweetalige skoolwoordeboek Afrikaans–Engels/Engels–Afrikaans*. In die voorwoord van hierdie woordeboek word die teikengebruikers bloot as "leerlinge van die middelbare skool" aangedui. Daar is geen vermelding van die huistaal van die teikengebruikers nie. Dit word ook pertinent gestel dat hierdie woordeboek 'n verkorte weergawe is van die *Tweetalige woordeboek*. Vanuit 'n hedendaagse leksikografiese perspektief is dit 'n veroordelende uitspraak want die leksikograaf gee daardeur te kenne dat die woordeboek nie as 'n skoolwoordeboek gekonseptualiseer is nie maar 'n blote verkorting van 'n woordeboek wat 'n heeltemal ander gebruikerstipe en daarom 'n andersoortige dataversameling en data-aanbod het. 'n Skoolwoordeboek mag naamlik nie net 'n verkorte weergawe van 'n omvattender woordeboek wees nie. As tipologiese kategorie in eie reg moet 'n skoolwoordeboek van meet af as sodanig beplan en saamgestel word. Een van die eerste tweetalige woordeboeke met Afrikaans en Engels, dié van Kritzinger et al. (1921), is as 'n skoolwoordeboek beplan en later aangepas en uitgebrei tot 'n omvattender woordeboek. Die moontlike sukses van die uitbreiding tot 'n ander tipologiese kategorie word nie hier bespreek nie. Ter sake is dat die woordeboek as skoolwoordeboek beplan en saamgestel is. Tweetalige woordeboeke is egter nie tipieserwys in hulle bewerking daarop gerig om aan die gebruikers 'n uitbreiding van kennis oor hulle huistaal te bied nie.

As woordeboek gerig op leerders binne die Suid-Afrikaanse taalomgewing kan die leksikograaf nie 'n gelyke kennis van die twee behandelde tale by 'n teikengebruiker verwag nie. Vir die onthalwe van die Afrikaanssprekende wat die afdeling Afrikaans–Engels gebruik om 'n Engelse vertaalekwivalent vir 'n Afrikaanse vorm te kry, sou verwag kon word dat leiding gegee moet word om die keuse van die korrekte vertaalekwivalent te verseker. Die leiding word op 'n uiters inkonsekwente manier gegee en ontbreek dikwels feitlik geheel en al. Vergelyk byvoorbeeld die artikel van die lemma *stamp* as werkwoord (in *Tweetalige skoolwoordeboek*) wat die volgende bewerking in die semantiese kommentaar bied:

knock, pound, hit, give a blow; stamp;
pound, crush; bruise; bump, jolt; thud;
pitch (of ship); ram (into a gun, throat);
fyn-, pound, crush, bray; ...

Om van 'n leerder te verwag om die juiste vertaalekwivalent vir 'n bepaalde optrede van *stamp* te kies, is uiters onbillik.

Ander tweetalige woordeboeke met Afrikaans en Engels, bv. dié van Pharos, stel dit glad nie duidelik op watter sprekersgroep die woordeboek gerig is nie. Daar kan egter aanvaar word dat die bewerking nie aan Afrikaanse huistaalsprekers 'n beter kennis van Afrikaans gaan gee nie. Dus val tweetalige skoolwoordeboeke buite die bestek van hierdie artikel.

Die fokus in hierdie artikel is op eentalige Afrikaanse woordeboeke en wel op dié wat primêr op gebruik deur huistaalsprekers gerig is. Ook in hierdie kategorie woordeboeke word daar meestal nie duidelik genoeg aangedui of die woordeboek op huistaalsprekers al dan nie gerig is nie. Die werklike gebruik-situasie van 'n bepaalde woordeboek loop ook nie noodwendig parallel met dit wat deur die leksikograaf beplan is nie. 'n Woordeboek soos die Afrikaanse *Nuwe woordeboek sonder grense* is saamgestel vir addisionele taalleerders van Afrikaans in grade 4 tot 6. Volgens die voorwoord is dié woordeboek geskik vir sowel eerste- as tweede-addisionele taalsprekers. Die werklikheid van die gebruik-situasie lewer egter bewys dat hierdie woordeboek ook dikwels in skole vir huistaalsprekers gebruik word. Dit is nogmaals 'n bevestiging van die reeds genoemde uitspraak van Prinsloo en De Schryver (Op koms) dat gebruikers dikwels 'n woordeboek van 'n laer vlak verkies.

In sommige woordeboeke wat nie in die titel, subtitel, voorwoord of in 'n ander buitetekst 'n aanduiding bied van die teikengebruikers nie, is die aard, omvang en daarstelling van die data ook sodanig dat 'n mens nie kan aflei op watter gebruikersgroep die woordeboek gerig is nie. Uit die titel van 'n woordeboek soos Educum-uitgewers se *Junior verklarende woordeboek* kan 'n mens aflei dat dit waarskynlik op laerskoolleerders gerig is maar daar is geen aanduiding of dit gerig is op huistaalsprekers al dan nie. Hierdie gebrekkige leiding en veral die gebrekkige beplanning onderliggend daaraan dra by tot die mindere sukses van so 'n woordeboek as 'n praktiese gebruiksinstrument in die hand van die leerder en as 'n taalleerinstrument in die hand van die leerkrag.

Vervolgens word aandag gegee aan verskillende Afrikaanse woordeboeke vir huistaalsprekers. 'n Kritiese bespreking word gevolg deur aanbevelings wat tot die verbetering van hierdie breë woordeboek-kategorie kan lei.

4. Afrikaanse skoolwoordeboeke

In die hieropvolgende afdelings word vyf woordeboeke wat gedeeltelik of ten volle ontwerp is vir gebruik in Suid-Afrikaanse skole as hulpmiddel vir huistaalleerders van Afrikaans aan die orde gestel. In sommige van hierdie woordeboeke word eksplisiet gemeld op watter grade hulle gerig is terwyl ander glad nie daarna verwys nie. Aangesien die woordeboeke nie almal op dieselfde graadgroep gerig is nie, word hulle in twee groepe verdeel wat afsonderlik bespreek word, te wete dié wat ontwerp is vir sommige grade in die laerskool en dié waarna daar in hierdie artikel op grond van bepaalde kenmerke verwys word as hoërskoolwoordeboeke, alhoewel die woordeboeke self geen aanduiding gee van hulle teikengroep nie.

4.1 Woordeboeke vir laerskoolgebruik

Die eerste woordeboek wat hier bespreek word, is Ben Conradie se *Verklarende Afrikaanse woordeboek vir laerskole* wat in 1967 verskyn het en een van die vroeë Afrikaanse woordeboeke is wat spesifiek vir skoolgebruik ontwerp is. Afgesien van die verwysing in die titel dat die woordeboek vir laerskole saamgestel is, word daar geen verdere inligting voorsien oor sy funksies of die grade waarvoor dit beplan is nie. Die kort inleidingstekste gee slegs te kenne dat dit in ooreenstemming met die jongste uitgawe van die *Afrikaanse Woordelys en Spelreëls* gemaak is. Die sentrale teks van die woordeboek beslaan 207 bladsye met ongeveer 12 000 lemmas, asook 'n agtertekstelys van 3 bladsye wat die algemeenste afkortings bevat. Die woordeboek het 'n tradisionele uitleg wat lyk soos die verstekwoordeboek vir volwassenes van daardie tyd. Die artikels bevat 'n lemma met 'n aanduiding van hoofklem en verbuiging, gevolg deur kort definisies wat soms deur voorbeeldsinne ondersteun word. In sommige gevalle, maar taamlik selde, word die woordsoort ook aangedui:

on'derstel (-le) (s.nw.), voorwerp waarop die boonste gedeelte rus

ondersteun', help, steun, bystaan

onderstreep', 'n streep onder iets trek; nadruk lê op (*ek onderstreep jou vermaning*)

Die dekades na die verskyning van hierdie woordeboek het min nuwe skoolwoordeboeke met 'n werklik vernuwend aanslag vir Afrikaanse huistaalleerders in die laerskool gelewer. 'n Volgende belangrike bydrae was die *Shuters Junior woordeboek* wat in 2001 verskyn het om aan Afrikaanse skoolleerders 'n leksikografiese produk te gee wat sterker gerig was op hulle spesifieke vlak en behoeftes. Soos sy voorganger verstrekte hierdie woordeboek min inligting oor homself en sy funksies. Die woord *junior* in die titel suggereer wel dat die woordeboek vir laerskoolleerders saamgestel is, en op die agterste buiteblad staan daar die woordeboek is "geskryf vir kinders vanaf 8 jaar oud" — dus waarskynlik vir leerders vanaf graad 2 of 3 en ouer. Die kort teks op die agterste buiteblad dui ook aan dat die woordeboek 3 300 woorde en 450 illustrasies bevat. Dit is "gebaseer op navorsing oor die woorde wat jong lesers die meeste benodig" en die "definisies gebruik hoëfrekwensiewoorde". Hiermee het die *Shuters Junior woordeboek* belangrike vooruitgang gemaak ten opsigte van 'n leksikografiese produk wat aangepas is by die spesifieke behoeftes van laerskoolleerders. Die aantal lemmas is omtrent een kwart van dit wat in die *Verklarende Afrikaanse woordeboek vir laerskole* aangebied word maar die lemmakeuse is gebaseer op wetenskaplike kriteria en navorsing oor die woordeskat wat gebruik en benodig word vir die teikengebruikersgroep. Daarbenewens is daar

nuwe beginsels ingevoer om die definisies verstaanbaarder te maak en hulle by die leerders se taalkundige en verstandelike ontwikkeling te laat aanpas.

Afgesien van die sentrale teks met sy hooflemmalys bevat die woordeboek ook 'n klein prentjietrajek in die agtertekstegedeelte met ses tematies georganiseerde afdelings (*Sport en ontspanning, Vervoer, Op die strand, Wonings, Skool en By die supermark*) waar gepaste woorde by die onderskeie prentillustorasies gevoeg is. Dit verteenwoordig 'n vernuwing, maar die werklike waarde kan bevraagteken word omdat dit slegs ses afdelings bevat — alhoewel daar in hierdie afdelings meer as 200 woorde met prentjies geïllustreer word. In vergelyking met die *Verklarende Afrikaanse woordeboek vir laerskole*, is die uitleg van die hooflemmalys ook 'n aansienlike verbetering. Hierdie afdeling beslaan 208 bladsye, amper dieselfde aantal as in die *Verklarende Afrikaanse woordeboek vir laerskole*, maar aangesien die nuwe woordeboek slegs omtrent 25 persent van die ouer woordeboek se aantal lemmas bevat, laat dit ruimte vir die gebruik van 'n groter en leesbaarder lettertipe asook vir meer ruimte tussen die artikels. Dit verhoog die toegang tot en leesbaarheid van die artikels. Daarbenewens word die letters van die alfabet op elke bladsy in vertikale rooi letters aangedui. Dit maak dit vir leerders makliker om die woorde te vind en help hulle ook met die bemeestering van die alfabet. Die artikels bestaan uit 'n lemma (in vet rooi letters), 'n kort definisie en, in omtrent 25 persent van die gevalle, 'n voorbeeldsin wat die gebruik van die woord illustreer en die definisie ondersteun. Die artikels gee nie woordsoortaanduiding nie en verbuiging word slegs in 'n redelik beperkte aantal gevalle verstrek waar daar ortografiese of ander probleme ondervind kan word. In die geval van polisemiese lemmas word die verskillende betekenisonderskeidinge deur syfers aangedui:

slaaf [slawe] iemand wat gedwing word om sonder betaling vir iemand anders te werk

slaag om reg te kry wat jy probeer het om te doen; om 'n toets of eksamen deur te kom *As jy hard werk, is ek seker dat jy jou eksamen sal slaag.*

slaai 1. 'n sappige soort groente wat uit groen blare bestaan **2.** 'n mengsel van vars groente of vrugte wat by 'n maaltyd voorgesit word

Die derde woordeboek wat hier bespreek word, is Maskew Miller Longman se *Grondslagfasewoordeboek Afrikaans/English* wat in 2010 verskyn het. Dit word duidelik gestel hierdie "woordeboek is geskik vir leerders in Graad 1 tot 4 en dek al die relevante assesseringstandaarde wat in die kurrikulum gespesifiseer word". Anders as in die geval van die voorafgaande twee woordeboeke gee

hierdie woordeboek duidelik leiding (p. iv) oor sy teikengebruikersgroep en sy funksies:

Hierdie woordeboek sal jou help om:

- praat-, lees- en skryfvaardighede in jou huistaal [Afrikaans] te ontwikkel
- Engels as Addisionele Taal te verstaan en te gebruik
- Afrikaans as Addisionele Taal aan te leer en te gebruik.

Ondanks hierdie inligting, asook die reeds vermelde verwysing na Engels in die titel, laat die data wat in hierdie woordeboek aangebied word, min twyfel dat dit sy hoofdoel is om Afrikaanssprekende skoolleerders te help met die ontwikkeling van hulle huistaal. Afgesien van die inhoudsopgawe, 'n baie didaktiese hulpvaardige gebruikersgids en 'n oop bladsy met slegs die alfabetletters op, waarop leerders hulle eie eenbladsywoordeboek kan maak, bestaan *Grondslagfasewoordeboek* uit:

- 'n 26-bladsy-afdeling, die sogenaamde "prentewoordeboek", met 18 tematiese trajekte wat ongeveer 290 van die ongeveer 700 woorde bevat wat in die woordeboek behandel word;
- die hooflemmalys (65 bladsye) waar al die bladsye 'n vertikale alfabet het en waar al die artikels klein illustrasieprente bevat; in die meeste gevalle met verwysing na die trajekte van die prentewoordeboek waar die prente in hulle "natuurlike" tematiese omgewing gevind kan word;
- twee tweetalige woordelyste, Afrikaans–Engels en Engels–Afrikaans, met kruisverwysings na die bladsye in die hooflemmalys waar die betrokke woord behandel word.

Die artikels in die sentrale teks bevat, naas die lemmas en prentillustrasies, verbuigings, Engelse ekwivalente, definisies, voorbeeldsinne in sowel Afrikaans as Engels en, waar beskikbaar, Afrikaanse sinonieme en antonieme. In sommige gevalle is daar ook teksskassies wat aan die lemma geadresseer is met opmerkings wat "meer inligting oor 'n woord en hoe om dit te gebruik" bevat. Die bladsy- en artikeluitleg met die benutting van 'n groot lettertipe en meer as genoeg ruimte tussen die artikels blyk goed aangepas te wees vir die geïdentifiseerde teikengroep. Hier volg drie artikels, maar sonder die prentillustrasies:

badkamer (badkamers) bathroom
'n Kamer waar jy jouself was, is 'n
badkamer.
*Die bad is in die badkamer./The bath is
in the bathroom.*

baie lots of, many
Baie is 'n groot aantal of 'n klomp —
nie min nie.

Daar is **baie** voëls in die boom./There are **lots of** birds in the tree.

Daar is **baie** bome in die park./There are **many** trees in the park.

TEENOOR: **min**

bak (bakke) dish

Iets waarin jy kos sit, is 'n **bak**.

Ons eet pap uit die **bak**. / We eat porridge from this **dish**.

Soos reeds genoem, lyk dit of die *Grondslagfasewoordeboek* goed aangepas is vir skoolleerders van graad 1–4. Die konsekwente gebruik van illustrasieprente en veral ook die tematiese prentwoordeboekafdeling maak dit moontlik om op verskillende maniere toegang tot die data in hierdie woordeboek te kry, selfs wanneer die gebruikers nog besig is om die alfabet te bemeester. Die gebruik van definisies in natuurlike taal wat die woord wat gedefinieer word, insluit, maak dit makliker vir kinders om die betekenis te verstaan. Ongelukkig is daar verskeie gevalle waar die definisies woorde bevat wat nie as lemmas opgeneem is nie en wat moeilik verstaanbaar vir skoolleerders kan wees. Die woord *familie* word soos volg verklaar:

'n Familie is 'n groot groep mense wat verwant is aan mekaar.

Nóg *mekaar* nóg *verwant* (wat moeiliker verstaanbaar vir grondslagfaseleerders kan wees as *familie*) is as lemmas opgeneem. Nog 'n voorbeeld is in die tekskassie met die lemma *lank* as adres waar daar staan:

NOTA: Kyk mooi hoe verander **lank** wanneer ons dit voor 'n selfstandige naamwoord gebruik ...

Geeneen van die woorde *verander*, *wanneer*, *selfstandig* en *naamwoord* word in die woordeboek verklaar nie en vir die woord *mooi* word slegs 'n ander betekenisonderskeiding verstrekk. Indien hierdie tipe fout reggestel word, sal die *Grondslagfasewoordeboek* nog oortuigender wees en 'n groot stap vorentoe in vergelyking met ander woordeboeke vir laerskoolleerders beteken, veral met betrekking tot die natuurliketaaldefinisies, funksionele uitleg, groot tematiese prentwoordeboekafdeling en die verskeidenheid toegangsroetes vir skoolleerders met verskillende tipes behoeftes en vlakke van taalkundige en verstandelike ontwikkeling. Daarbenewens het hierdie woordeboek dalk selfs twee bykomende funksies wat nie genoem word nie, naamlik om regstreeks te help met woordeskatuitbreiding deur middel van die tematiese prentwoordeboekafdeling en om skoolleerders in te lei in woordeboekgebruik.

4.2 Woordeboeke vir die hoërskool

In die Afrikaanse leksikografie het 'n woordeboek vir huistaalleerders van Afrikaans op hoërskoolvlak lank ontbreek. 'n Vernuwend bydrae in hierdie verband was die verskyning van F.F. Odendal se *Kernwoordeboek* in 2000. Alhoewel hierdie woordeboek geen aanduiding gee van sy teikengebruikersgroep nie, plaas persoonlike opmerkings deur die skrywer dit in die kategorie van 'n skoolwoordeboek en uit die inhoud kan afgelei word dat dit primêr op hoërskoolleerders gerig is. In die "Voorwoord" sê die skrywer dat die woordeboek beplan is as 'n "verklarende woordeboek", en hy benadruk:

Die strewe was om die kernwoordeskat van Afrikaans te verklaar.

In hierdie opsig is die hoof verklaarde funksie van hierdie woordeboek om gebruikers te help met teksresepsie. Dit is egter nie heeltemal duidelik of dit ook bedoel is om die gebruiker met teksproduksie te help nie, alhoewel Odendal in die voorwoord skryf:

Daar is verder vryelik van voorbeeldsinne gebruik gemaak want dit is ten slotte wat die gebruiker van 'n woordeboek wil weet: Hoe kan ek die woord werklik *gebruik* nou dat ek weet wat sy betekenis is?

'n Vinnige berekening toon dat die *Kernwoordeboek* ongeveer 15 000 artikels bevat wat op elke bladsy in drie kolomme gerangskik is. Hierdie rangskikking gee aan die woordeboek 'n baie kompakte voorkoms wat ongetwyfeld toegang tot die verlangde data bemoeilik. Die artikels bestaan uit 'n lemma (met aanduiding van lettergreepverdeling, hoofklem, woordsoort, verbuiging en verkleiningsvorm) gevolg deur 'n definisie en voorbeeldsinne. Sowel die definisie as die voorbeeldsinne sluit die woord in wat deur die lemma verteenwoordig word. Aan die einde van sommige artikels is daar ook vaste uitdrukkings ("idiome, spreekwoorde, gesegdes, ens.") en relevante kruisverwysings na ander artikels. In die geval van polisemiese lemmas word die verskillende polisemiese onderskeidings elk deur 'n syfer as polisemiemerker voorafgegaan. Aan sommige artikels is daar ook 'n lemmanis geheg met onverklaarde samestellings en afleidings:

frus·tra'sie *s.nw.* (frustrasies; frustrasietjie)
'n Frustrasie is iets wat 'n mens verhinder om te bereik wat jy wil doen:
Klein frustrasies oorkom. Haar negatiewe houding was 'n groot frustrasie.
frustreer'.

fuif *ww.* (gefuif)
As 'n mens fuif, neem jy deel aan 'n jolige eet-en-drink-party: *Tot die oggend*

vroeg fuif.

fuif'party. fui'wer(y).

fun·geer' *ww.* (gefungeer)

As 'n mens fungeer, neem jy waar in 'n posisie, verrig jy die diens of werksaamheid van iemand: *As skoolhoof fungeer vir een kwartaal. Die dokter fungeer sommer ook as veearts.*

funk'sie *s.nw.* (funksies)

1 Die taak, werk wat iemand of iets moet verrig, is hulle funksie: *'n Belangrike funksie binne die maatskappy hê. Die funksie van die hart is om die bloed in omloop te hou. Die nuwe stuk masjinerie verrig sy funksie uitstekend.*

2 'n Funksie is 'n sosiale, feestelike geleentheid: *Die rugbyklub se jaarlikse funksie.*

Die voorafgaande artikels bevat ook data, byvoorbeeld lettergreepverdeling, hoofklem, verkleiningsvorme, afleidings, ens. wat die gebruikers nie met teksresepsie nie maar wel met teksproduksie ondersteun. Indien die woordeboek regtig beplan is om skoolleerders met teksproduksie te help, sou 'n mens verwag het om hierdie bykomende data te vind maar veral ook data met betrekking tot sintaktiese kenmerke, kollokasies, sinonieme en antonieme. In hierdie opsig word daar slegs gedeeltelik voorsien in die teksproduksiefunksie. Maar dit is sonder meer duidelik dat die bogenoemde artikels wel hulp verleen met teksresepsie. Dit is ongetwyfeld so dat die samesteller deeglik nagedink het oor die beste manier waarop die definisies aangebied moet word. Hy skryf soos volg daaroor in die voorwoord:

Die metode wat nog slegs deur enkele moderne woordeboeke gebruik word, is toegepas, naamlik om die definisie in 'n volsin te gee waarin die hoofwoord self voorkom. Daardeur lees die definisie nie net makliker nie, en is dus nie net makliker verstaanbaar nie, maar die hoofwoord word onmiddellik gegee in 'n verband waarin dit in gewone Afrikaans gebruik word.

Ondanks hierdie tersaaklike stelling moet 'n mens dit tog wel oorweeg om na te gaan of al die definisies ewe suksesvol aangepas is by die taalkundige, verstandelike en kulturele vlak van die teikengroep en of, selfs deur dieselfde definiëringsbeginsels toe te pas, die definisies in 'n makliker styl geskryf kon gewees het, gebaseer op 'n beperkte definiëringswoordeskat. Eweneens kan ook gevra word of die woordeboek werklik die *kernwoorde* bevat wat hoërskoolleerders nodig het en of ander beginsels gebruik moes gewees het in die lemmakeuse om die lemmalys aan te pas by die woorde waarmee die leerders op skool te doen kry en wat hulle veronderstel is om te beheers.

Die tweede woordeboek wat in hierdie afdeling bespreek word, is die *HAT Afrikaanse skoolwoordeboek* wat in 2009 gepubliseer is. Alhoewel hierdie woordeboek 'n onafhanklike publikasie in eie reg is, is dit baie aan die *Kernwoordeboek* as voorganger verskuldig. Hierdie erkenning word duidelik in die voorwoord deur die redakteur, Jana Luther, gegee:

Benewens die redaksionele span is ek groot erkentlikheid verskuldig aan prof. F.F. Odendal, wie se *Kernwoordeboek* van 2000 as basis vir die nuwe woordeboek gebruik is. Sonder dié stewige grondslag waarop ons kon voortbou, sou die publikasie van hierdie woordeboek vanjaar onmoontlik gewees het.

Die *HAT Afrikaanse skoolwoordeboek* toon baie verbeteringe ten opsigte van sy voorganger maar ook enkele terugwaartse treë. Een aspek waar die status quo gehandhaaf is, is die vlak van die inligting. Behalwe deur te sê dat dit 'n skoolwoordeboek is wat gebruik moet word deur "Afrikaans Huistaal- en Afrikaans Eerste Addisionele Taalleerders" en "elke gebruiker van Afrikaans, nie net leerders nie", word daar niks gesê oor die grade van die teikengebruikers of die funksies van die woordeboek nie. Uit die inhoud wat hierna bespreek word, blyk dit dat hierdie woordeboek primêr vir hoërskoolleerders ontwerp is en nie ewe suksesvol deur laerskoolleerders gebruik sal kan word nie.

Afgesien van die voorwoord en die sentrale teks met die hooflemmalys bevat die *HAT Afrikaanse skoolwoordeboek* 'n kort sleutel tot die woordeboek asook 'n lys redaksionele afkortings, aangebied in die voorteksteafdeling, en, in die agterteksteafdeling, 'n afkortingslys en 'n lys "belangrike terme vir Afrikaans Huistaal en Afrikaans Eerste Addisionele Taal" — 'n bondige geïntegreerde woordeboekie in eie reg. Op die binnekant van die voorste buiteblad is 'n kort lysie intensiewe vorme en op die binnekant van die agterste buiteblad 'n lys vergelykings. Aangesien die gebruikersgids niks oor hierdie twee tekste sê nie, is dit moeilik om die rede vir hulle opname te bepaal.

Die aanbod in die sentrale teks is baie kompleks as dié van enige van die reeds bespreekte woordeboeke. Volgens die inligting op die agterste buiteblad bevat hierdie woordeboek 30 000 woorde en 300 illustrasies maar volgens 'n vinnige berekening lyk dit eerder na 15 000 artikels. Dit is nie duidelik hoe daar by 30 000 uitgekom is nie, behalwe as afgeleide vorme wat as sublemmas opgeneem is, ook daarby ingesluit is. Die artikels bestaan telkens uit 'n lemma met 'n aanduiding van hoofklem, lettergreepverdeling, verbuiging, verkleiningsvorm en 'n definisie wat meestal deur 'n voorbeeldsin ondersteun word. Anders as in die *Kernwoordeboek* word deeltjiewerkwoorde in hulle sintetiese en analitiese vorm aangebied. Dit is 'n welkome toevoeging. Soms verskaf die artikels ook data oor sintaktiese kenmerke, idiomatiese uitdrukkings, spreekwoorde, sinonieme, antonieme en etimologie. Partykeer word 'n artikel gevolg deur 'n tekskassie met inligting oor woorde met 'n soortgelyke betekenis (tesouruskassies), gebruik, spelling en skryfwyse:

aan·leg *s.nw.* [*~te of ~gings of ~ginge of aanlêe*] **1** 'n plek wat vir 'n spesiale doel ontwerp is: 'n *aanleg om olie uit steenkool te verkry*. **2** die talent wat iemand het: 'n *aanleg vir kuns/musiek/wiskunde hê*.

aan·lei·ding *s.nw.* [*~s of ~e*] **die aanleiding tot iets** die rede waarom iets gebeur: *die aanleiding tot 'n oorlog/rusie/opstand*. | **na aanleiding van ...** *Na aanleiding van u vraag/brief antwoord ek graag soos volg*.

aan·lok *ww.* [**lok aan; het aangelok iem./iets aanlok** iets aangenaams sê/voorstel/wys/aanbied om iemand/iets na jou toe of na 'n plek te laat kom: *kliënte aanlok met advertensies* ° *Bye word deur die geur van blomme aangelok*.

■ **aan·lok·lik** *b.nw., bw.* [*~e; aanlokliker, die aanloklikste*] baie aantreklik of verleidelik; wat iemand/iets aantrek: 'n *aanloklike aanbod/prys* ° 'n *aanloklike kampeerplek*.

TESOURUS

aantreklik 'n *aantreklike werksomgewing*.

begeerlik 'n *begeerlike skildery*.

onweerstandbaar *onweerstandbare geure*.

verleidelik *verleidelike perskes*.

aan·loop ■ *ww.* [**loop aan; het aange-loop**]

1 verder loop: *Rus julle maar 'n bietjie; ek loop verder aan*. ° *As jy 'n entjie aanloop, sal jy die winkel op die hoek kry*. **2** vinniger loop: *Ons moet aanloop, anders is ons laat*.

■ *s.nw.* [**aanlope**] **1** 'n klomp mense wat iewers aankom: *Daar is altyd 'n hele aanloop van vriende by die meisie se huis*. **2** die inleiding tot iets: *Die aanloop tot sy toespraak was maar verovelig*. **3** die afstand/beweging voordat 'n handeling uitgevoer word: 'n *bouler/hoogspringer se aanloop*.

- **aan-loop-baan** s.nw. [aanloop-bane] die strook waarvandaan vliegtuie opstyg en waarop hulle land.

Soos uit hierdie voorbeeld blyk, bevat die *HAT Afrikaanse skoolwoordeboek* die meeste data wat nodig is om gebruikers met sowel teksresepsie as teksproduksie te help. Dit is beduidende vordering vanaf die *Kernwoordeboek* waarop dit gedeeltelik gebaseer is. Die aanbod is egter nie oral op die beste moontlike en mees gebruikersvriendelike manier nie. Dit kan egter in 'n tweede uitgawe van die woordeboek verbeter word. In die hieropvolgende paragrawe word vyf van die belangrikste probleme in hierdie woordeboek bespreek.

Die eerste probleem blyk uit die tradisionele definisiemethode met sy bondige betekenisverklarings wat die *Kernwoordeboek* se metode om natuurlike-taaldefinisies te gebruik, vervang het. Onses insiens bied hierdie verandering nie die beste oplossing vir hoërskoolleerders nie. Die definiëringsmetodes wat deur sy voorganger ontwikkel is, moes liefers in hierdie woordeboek gehandhaaf en selfs verder ontwikkel gewees het.

'n Tweede probleem wat ons ook as 'n terugwaartse stap in vergelyking met die *Kernwoordeboek* beskou, is die gereelde aanduiding van twee of meer verbuigings- of spellingvariante, soos blyk uit die bogenoemde artikels van die lemmas *aanleg* en *aanleiding*. Hierdeur word die variasie wat deur die Taalkommissie in sy *Afrikaanse woordelys en spelreëls* erken word, wel weerspieël maar die vraag is of die manier waarop dit hier aangebied word vir die gebruiker die beste moontlike hulp verleen. Gebruikers is onseker watter een om te kies wat die raadplegingstyd verleng. Hulle kies dan meestal op 'n lukraak manier en die keuse is dikwels ten gunste van 'n minder frekwente of minder bekende vorm. In 'n woordeboek vir volwassenes sou hierdie deskriptiewe prosedure beter kon werk maar in 'n skoolwoordeboek behoort 'n proskriptiewe metode eerder gevolg te word met die leksikograaf wat slegs een variant gee of meerdere variante aanbied maar een bepaalde vorm aanbeveel (vgl. Tarp en Gouws 2008).

Die derde probleem spruit uit die kompaktheid van die woordeboekartikels. Op elke bladsy is die deeltrajekte in twee kolomme gerangskik. Dit is 'n verbetering op die drie kolomme in die *Kernwoordeboek* maar as gevolg van al die data in elke artikel het die artikels 'n uiters verdigte en kompakte voorkoms wat kits-toegang tot die data kan benadeel en ook die leesbaarheid en daarmee die werklike bruikbaarheid van die woordeboek in gedrang kan bring. Verskeie aspekte dra tot hierdie probleem by, byvoorbeeld dat afgeleide vorme wat as verklaarde vorme opgeneem word, as sublemmas van die artikel van die basisvorm, die hooflemma, behandel word, soos blyk uit die artikels *aanlok* en *aanloop* waar *aanloklik* en *aanloopbaan* onderskeidelik as verklaarde sublemmas gegee word. Sowel *aanloklik* as *aanloopbaan* kon eerder as onafhanklike hooflemmas opgeneem gewees het aangesien elk van hulle die adres is van 'n

verskeidenheid aanduiders wat die verstekdataoordrag van 'n artikel bewerkstellig. Die tradisie wat in talle woordeboeke vir volwassenes gevolg word, naamlik om grammatiese homonieme in dieselfde artikel te plaas, word ook hier gevolg. Die vraag kan gevra word of dit die aangewese aanbiedingswyse vir 'n skoolwoordeboek is. Indien sulke vorme in verskillende artikels sou optree, kon dit toegang tot hierdie vorme vergemaklik het en die artikels minder kompak gemaak het. Alhoewel redaksionele en ander afkortings in onderskeie buitetekste in die voor- en agterteksteafdelings aangebied word, dra die veelvuldige afkortinggebruik in die woordeboekartikels wat die gebruiker nie altyd met die ooreenstemmende volvorme in verband kan bring nie, by tot die kompakteit van die artikels. Die gebrek aan 'n gebruikersvriendelike artikel-interne soekveldstruktuur en mikroargitektuur (vgl. Bergenholtz, Tarp en Wiegand 1999) maak dit moeilik om vinnig te kan onderskei tussen die aanduiders wat die verskillende datatipes oordra. Die feit dat die verklarings van elke nuwe betekenisonderskeiding nie soos in die *Kernwoordeboek* op 'n nuwe reël begin nie, verhoog ook die kompakteit van die artikels. Met 'n gebruikersvriendeliker mikroargitektuur en soekveldstruktuur kon die artikel van die lemma *aanloop* soos volg gelyk het:

aan·loop ww. [loop aan; het aange-loop]

1 verder loop: *Rus julle maar 'n bietjie; ek loop verder aan.* ◦ *As jy 'n entjie aanloop, sal jy die winkel op die hoek kry.*

2 vinniger loop: *Ons moet aanloop, anders is ons laat.*

aan·loop s.nw. [aanlope]

1 'n klomp mense wat iewers aankom: *Daar is altyd 'n hele aanloop van vriende by die meisie se huis.*

2 die inleiding tot iets: *Die aanloop tot sy toespraak was maar vervelig.*

3 die afstand/beweging voordat 'n handeling uitgevoer word: *'n bouler/hoogspringer se aanloop.*

aan·loop·baan s.nw. [aanloopbane]

die strook waarvandaan vliegtuie opstyg en waarop hulle land.

'n Vierde probleem is die hantering van homonieme in die *HAT Afrikaanse skoolwoordeboek*. Hulle word op dieselfde manier aangebied as in baie Afrikaanse en ander woordeboeke. Dit is egter 'n geval waar die eiesoortige behoefte van die leerder nie voldoende in aanmerking geneem is nie. Die linguistiese kriteria vir die vasstelling van homonieme lei nie altyd tot 'n ondubbel-

sinnige klassifikasie nie. As dit 'n probleem vir taalkundiges en leksikograwe is, is dit des te moeiliker vir die teikengebruikers van 'n skoolwoordeboek om tussen homonieme te onderskei en om dan in 'n woordeboek vas te stel onder watter lemma 'n bepaalde betekenisverklaring gesoek moet word. Vergelyk die opname en bewerking van die twee homonieme *klink*:

klink¹ *ww.* [**het ge~**] 1 'n helder geluid
maak: ... 2 die genoemde indruk wek
wanneer jy iets hoor/lees ... 3 **op**
iem. / iets klink ...

klink² *ww.* [**het ge~**] iets met 'n hamer
of ander slaanding plat slaan, of
onderdele met klinknaels (= metaal-
pennetjies) vassit ...

Vir die tipiese teikengebruiker van hierdie woordeboek is die verskynsel van homonimie wat hier gemerk word, nie noodwendig belangrik nie. Belangrik is dit wel om die betekenis van die woord *klink* wat hy/sy iewers teëgekomp het, te vind. Dié gebruiker word nie gehelp in sy teksresepsieprobleem deur die afsonderlike artikels nie. Vir teksresepsie- en teksproduksiebehoefes is dit nie belangrik vir 'n leerder om te weet of twee woorde homonieme is en of twee betekeniswaardes betekenisonderskeidings van 'n enkele polisemiese leksikale item is nie. Die leerder moet die woord in die woordeboek vind en dan die tersaaklike betekenisverklaring of ander verstrekte inligting kan onttrek. Hier is die plasing van homonieme volgens linguïstiese beginsels in afsonderlike artikels minder ter sake. Een lemma kan eerder as gidselement optree wat die gebruiker na die tersaaklike data-aanbod lei. Vergelyk in hierdie verband ook Tarp (2009).

Die vyfde probleem het te make met die aanduiding van sintaktiese kenmerke alhoewel, soos reeds genoem, die *HAT Afrikaanse skoolwoordeboek* in hierdie verband 'n groot verbetering op sy voorganger, die *Kernwoordeboek*, toon. Ons bespreek nie die waarde al dan nie van die insluiting van nog meer relevante sintaktiese data nie maar slegs die manier waarop hierdie data aangebied is — 'n werkswyse wat onses insiens (nogmaals) tot 'n te kompakte aanbod lei wat hulp aan die spesifieke teikengebruikersgroep verminder. Vergelyk byvoorbeeld weer die aanbieding, plasing en bewerking van die inskrywings *iem. / iets aanlok* in die artikel van die lemma *aanlok*:

aan·lok *ww.* [**lok aan; het aangelok**]
iem. / iets aanlok iets aangenaams sê/
voorstel/wys/aanbied om iemand/
iets na jou toe of na 'n plek te laat
kom: *kliënte aanlok met advertensies* ◦
Bye word deur die geur van blomme aangelok.

Volgens die sleutel tot dié woordeboek word *iem./iets aanlok* hier as woord-groepe aangebied. Hierdie inskrywings word voor die betekenisverklaring geplaas en word dus die logiese adres van dié verklaring in plaas daarvan dat die lemmatiese adressering duidelik blyk. In hierdie artikel is *iem./iets aanlok* twee verdigte inskrywings wat eksplisiete sintaktiese data bied. In stede van hierdie plasing en aanbod kan hierdie inskrywings in 'n minder verdigte vorm as *iemand aanlok* en *iets aanlok* na die betekenisverklaring volg en wel elkeen in 'n afsonderlike reël begin. Dit is 'n gebruikersvriendeliker soekveldstruktuur en mikroargitektuur. Die bewerking van hierdie inskrywings kan dan 'n eie voor-beeldsin insluit ter illustrasie van die optrede van so 'n sintaktiese konstruksie. Hierdeur sou die aanbieding aansluit by beginsels wat voorgestel is in Tarp en Gouws (2004) met betrekking tot woordeboeke vir niemoedertaalleerders van Afrikaans, 'n gebruikersgroep wat volgens die redakteur deel vorm van die teikengroep van die *HAT Afrikaanse skoolwoordeboek*. Die toepassing van hierdie werkswyse sou tot die volgende aanbod kon lei:

aan·lok *ww.* [**lok aan; het aangelok**]

iets aangenaams sê/voorstel/wys/
aanbied om iemand/iets na jou toe of
na 'n plek te laat kom: *kliënte aanlok
met advertensies.* ◦ *Bye word deur die
geur van blomme aangelok.*

iemand aanlok: *Met sy mooi woorde
het hy my na sy huis aangelok.*

iets aanlok: *Die fluitspeler se musiek
het al die rotte in die dorp na hom aan-
gelok.*

Die voorafgaande opmerkings kan slegs gemaak word omdat daar soveel verbeteringe in die *HAT Afrikaanse skoolwoordeboek* aangebring is in vergelyking met sy voorganger, die *Kernwoordeboek*, waaraan hy volgens die redakteur 'n groot deel van sy bestaan te danke het.

5. Aanbevelings

Uit die voorafgaande bespreking is dit duidelik dat skoolwoordeboeke 'n belangrike tipologiese klas verteenwoordig en veel daartoe kan bydra om 'n woordeboekgebruikskultuur te vestig. Alhoewel hierdie bespreking primêr gerig is op skoolwoordeboeke vir huistaalsprekers van Afrikaans, is die bevindinge wat ten opsigte van die Afrikaanse woordeboeke gemaak is asook die volgende aanbevelings nie uitsluitlik op Afrikaans gerig nie. Die leksikografieteorie het 'n nietaalspesifieke benadering en dit geld ook die toepaslikheid van opmerkings wat ten opsigte van woordeboeke van 'n spesifieke taal gemaak word.

Die vyf woordeboeke wat in hierdie bespreking behandel is, lewer ruimskoots daarvan bewys dat hedendaagse skoolwoordeboeke, vergelyk byvoorbeeld die *Grondslagfasewoordeboek* en die *HAT Afrikaanse skoolwoordeboek*, daarop ingestel is om aan bepaalde gebruikersbehoefes te beantwoord. Skoolwoordeboeke toon 'n duidelike invloed van die teoretiese leksikografie en hierdie invloed dra kennelik daartoe by dat die gehalte van die woordeboeke verbeter het. Die wisselwerking tussen teorie en praktyk in die leksikografie kan egter nog veel verder gevoer word om tot nog beter leksikografiese produkte te lei. Wat skoolwoordeboeke betref, is dit belangrik dat uitgewers, praktiserende leksikograwe, metaleksikograwe en opvoedkundiges as 'n span moet saamwerk om die beste moontlike skoolwoordeboeke vir die onderskeie teikengebruikersgroepe saam te stel. In hierdie verband is dit belangrik dat inligting verkry moet word vanaf die opvoedkundiges, veral opvoeders, maar ook leerders wat met die betrokke woordeboeke moet werk. Hier mag die intuïsie van die leksikograaf of uitgewer nie die enigste maatstaf wees nie. Empiriese gebruikersnavorsing is noodsaaklik.

Wat ook belangrik is, is dat enige woordeboek wat tot die kategorie van skoolwoordeboeke behoort, liefs nie in isolasie beplan moet word nie. 'n Uitgewer wat onderneem om 'n skoolwoordeboek uit te gee — al is dit slegs 'n enkele skoolwoordeboek vir slegs een van die fases in die omvattender skoolprogram — behoort so 'n woordeboek as deel van 'n subtipologiese kontinuum te beplan. In hierdie beplanningsproses behoort die volle omvang van die skoolwoordeboekbehoefes ondersoek te word en gekyk te word hoe elk van die verskillende subtypes skoolwoordeboeke 'n eie gedeelte van die kontinuum moet dek. Die leksikograaf moet uiteindelik weet waar pas 'n bepaalde skoolwoordeboek binne hierdie kontinuum en hoe hou so 'n woordeboek verband met die ander woordeboeke wat ook in dieselfde kontinuum sou pas. Naas leksikografiese beginsels, onder meer kwessies soos die funksies, strukture en inhoud van woordeboeke, moet opvoedkundige beginsels ook 'n rol speel. In hierdie verband is dit belangrik dat die bepaling van die relevante nasionale skoolkurrikula in ag geneem moet word. Waar woordeboekgebruik deur kurrikula vereis word, moet skoolwoordeboeke vir die betrokke leerdergroep so saamgestel word dat die tersaaklike woordeboekgebruiksvaardighede deur die raadpleging van die betrokke woordeboek ontwikkel kan word.

Die bespreking van die genoemde woordeboeke het duidelik getoon dat daar te min aandag gegee word aan 'n ondubbelsinnige spesifisering van die teikengebruikers vir wie 'n bepaalde woordeboek saamgestel word. Die kategorie "skoolwoordeboek" moet as 'n superordinaatkategorie gesien word wat in tipologiese hiponimiese verhouding tot 'n verskeidenheid subordinaatklasse staan. Om slegs die superordinaatklas te identifiseer, is nie voldoende nie. Leksikograwe en uitgewers moet veel presieser as dit wees. Elke skoolwoordeboek moet eksplisiete leiding bied, liefs reeds op die buiteblad aangedui, wat aantoon vir watter gebruikerssituasie en vir watter gebruikersgroep die woordeboek saamgestel is.

In hierdie nadere spesifisering speel bepalings soos die skoolgraad, ouderdom, kenmerke van die gebruikersgroep (of hulle huistaalsprekers of nie-moedertaalsprekers van die betrokke behandelde taal is), die gebruiksituasie en daarom ook die leksikografiese funksies van die beplande woordeboek 'n belangrike rol. Wat die funksies betref, behoort dit duidelik te wees dat daar onderlinge verskille tussen woordeboeke vir verskillende grade sal wees. In die grondslagfase is die aanleer van woordeskat, dit is ter vervulling van 'n kognitiewe funksie, aanvanklik belangrik. Later speel teksresepsie en teksproduksie ook 'n rol. In die gevorderde grade is teksproduksie toenemend belangrik maar teksresepsie bly ter sake terwyl die kognitiewe funksie 'n mindere rol speel in algemene woordeboeke. By die beplanning van 'n woordeboek moet die tersaaklike funksies reeds geïdentifiseer en die woordeboek dienoreenkomsig saamgestel word. 'n Deeglike beplanning in hierdie verband asook 'n aanduiding van die teikengebruikersgroep en hulle gebruiksituasie is nie net tot voordeel van die uiteindelijke of potensiële gebruikers van die woordeboek nie. Dit is grootliks tot voordeel van die leksikograaf wat die woordeboek moet opstel, aangesien duidelikheid oor hierdie kenmerke hom/haar help om 'n woordeboek tot stand te bring wat 'n reaksie is op spesifieke behoeftes van spesifieke gebruikers in spesifieke gebruiksituasies.

Daar bestaan sowel ruimte vir as behoefte aan graadspeesifieke Afrikaanse woordeboeke vir moedertaalsprekers. Die voorstelle wat hier gemaak word, veronderstel nie dat die verskillende hier genoemde subtypes gelyktydig gepubliseer moet word nie. Uitgewers, opvoedkundiges en leksikograwe moet egter verdere verlore woordeboekgebruikergenerasies probeer voorkom. Om dit moontlik te maak, kan nuwe skoolwoordeboeke ter aanvulling van die bestaande woordeboeke algaande ingefaseer word. Hier speel kennis van die reeds genoemde leksikografiese kontinuum 'n bepalende rol en elke nuwe woordeboek moet daarvolgens beplan en saamgestel word. Vier verskillende woordeboeke word voorgestel om voorsiening te maak vir die omvang en verskeidenheid van die skooljare. Hierdie woordeboeke is vir graad 1-3, graad 4-7, graad 8-10 en graad 11-12. Die woordeboek vir graad 11- en 12-leerders sal ook na skool in die tersiêre opvoeding of selfs as algemene woordeboek bruikbaar wees.

Dit is belangrik dat hierdie woordeboeke telkens so saamgestel moet word dat hulle reeds vir leerders in die laagste graad van die betrokke fase bruikbaar is. Uiteraard sal daar ook meer gevorderde inskrywings wees wat die senior gebruikers ter wille is maar die geïdentifiseerde behoeftes van die leerders van elk van die grade waarvoor 'n woordeboek saamgestel is, moet bevredig kan word. Selfs data wat ter wille van die senior gebruikers ingesluit is, mag nie van so 'n aard wees dat dit die junior gebruikers se woordeboekraadplegingsprosesse benadeel nie. Die verstek moet die behoeftes van die leerders in die eerste van die betrokke grade wees. Deur verrykingsinskrywings kan data vir die meer senior leerders aangebied word. In 'n grondslagfasewoordeboek het die graad 1-leerders moontlik nog nie 'n behoefte aan kollokasies nie terwyl die

graad 3-leerders dit wel het. Kollokasies kan as sintaktiese verrykingsinskrywings aangebied word maar wel op so 'n manier dat die graad 1-leerder se soektogte nie daardeur benadeel word nie. Die aard en omvang van die definisies word eweneens deur die gebruiker bepaal. Vir die gebruikers van 'n woordeboek vir graad 11- en 12-leerders is volsindefinisies dalk nie so belangrik om suksesvolle teksresepsie moontlik te maak nie; vir die leerders van laer grade is dit wel belangrik.

Hausmann en Wiegand (1989) het reeds aangetoon dat 'n woordeboek uit minstens twee tekste moet bestaan, te wete die sentrale teks wat die hooflemmalys en artikeltrajekte bevat, en minstens een buitetekst, gewoonlik in die voorteksteafdeling aangebied, wat 'n gebruikershandleiding vir die spesifieke woordeboek bevat. Vanuit die teorie van leksikografiese funksies word daar twee tipes data in 'n woordeboek verskaf, te wete primêre (funksieverwante) en sekondêre (gebruiksverwante) data. Die gebruikershandleiding is deel van hierdie sekondêre data. Te dikwels word die waarde van hierdie teks misken. Vir suksesvolle woordeboekgebruik is dit noodsaaklik dat ook die data wat hier aangebied word en die manier waarop dit aangebied word, die resultaat moet wees van deeglike beplanning, met inagneming van die teikengebruikersprofiel. Geen woordeboek behoort vandag sonder 'n gebruikershandleidingstekst te verskyn nie. In neskoolwoordeboeke bied so 'n handleiding 'n teks wat aan die teikengebruiker leiding gee oor wat in die betrokke woordeboek aangebied word en hoe die aangebode data ten beste gevind kan word. Dit moet ook in skoolwoordeboeke gedoen word maar hier moet die vlak van die leerders wat teikengebruikers is, in aanmerking geneem word. Die uiteensetting van die teks moet in ooreenstemming met die teksresepsievermoë van hierdie gebruikers wees. In skoolwoordeboeke het die gebruikershandleidingstekste egter ook 'n verdere rol. 'n Skoolwoordeboek is nie net 'n instrument in die hande van die leerder nie maar ook in die hande van die opvoeder. Die gebruikershandleiding kan 'n afdeling bevat waarin die opvoeder gehelp word met raad oor hoe die betrokke woordeboek ten beste as onderrigmiddel aangewend kan word. Opvallend in die bespreekte woordeboeke is die verskil in die aard en omvang van gebruikershandleidingstekste. Dit wissel van uiters beperk tot voldoende. In die toekoms behoort daar in die beplanning en samestelling van skoolwoordeboeke veel meer erns aan hierdie tekstipes bestee te word.

Die datakeuse in enige woordeboek moet in ooreenstemming met die teikengebruikers se behoeftes gedoen word. Dit geld die data in die buitetekste, die lemmaversameling asook die mikrostrukturele aanbod. In skoolwoordeboeke sal daar 'n beduidende verskil tussen die lemmakeuse vir grondslagfaseleerders en dié vir senior hoërskoolleerders wees. Skoolwoordeboeke se woordeboekkonseptualiseringplan en die daaropvolgende samestelling van die woordeboek moet dus kriteria in ag neem wat spesifiek vir hierdie lemma- en datakeuse ontwerp is. Toegang tot korpusse en die klem wat dikwels op gebruiksfrekwensie as criterium vir die lemmakeuse geplaas word, speel ook hier 'n rol maar leksikograwe moet nie gebruiksfrekwensie soos uit 'n algemene

korpus blyk sonder meer as enigste kriterium gebruik nie. Skoolwoordeboeke se lemmakeuse moet liefers volgens 'n toegewysde korpus gedoen word wat saamgestel is uit tekste wat dié materiaal bevat waarmee die betrokke leerders daaglik in hulle skoolleerplanne in aanraking kom. Sulke woorde wat miskien nie tot die algemene lys van hoëfrekwensiewoorde behoort nie, moet gebruik word om die korpus se hoëfrekwensiewoorde aan te vul. Ook in die lemma-keuse speel die beplande funksies van 'n woordeboek 'n wesenlike rol. Waar teksresepsie die beplande funksie is, moet 'n woordeboek 'n uitgebreider lemmaversameling hê as wanneer die klem op teksproduksie val. Teksresepsie-behoefte vir leerders strek dikwels verder as die tipiese hoëgebruiksfrekwensiewoorde van die betrokke graad- of ouderdomsgroep. Die resultate van 'n korpus se frekwensieaanduiding is nie genoeg motivering vir die lemmakeuse nie. In die beplanning van die reeds genoemde Afrikaanse woordeboek vir addisionele taalsprekers *Nuwe woordeboek sonder grense* is 'n toegewysde korpus gebruik wat uitsluitlik uit data saamgestel is wat in handboeke en studiemateriaal van die betrokke graadgroepe voorkom. Hierdie korpus is teenoor 'n algemene korpus gestel en aangevul met hoëgebruiksfrekwensiewoorde uit daardie korpus. Die gekombineerde korpus is vervolgens as basis gebruik vir die lemmakeuse in hierdie woordeboek. 'n Soortgelyke benadering kan ook in skoolwoordeboeke vir moedertaalsprekers gevolg word.

Die mees tipiese inligting wat gebruikers aan 'n eentalige woordeboek wil onttrek, is geleë in die data wat die betekenisverklaring verstrekk. Die manier waarop hierdie data in die leksikografiese definisie of betekenisparafrase aangebied word, is van bepalende belang vir die sukses van die betrokke woordeboek. In die ontwikkeling van die leksikografiepraktyk was die verskyning van die eerste uitgawe van die *Collins COBUILD English Language Dictionary* in 1987 van baanbrekende belang, onder meer vanweë die manier waarop die betekenisverklaring in hierdie woordeboek aangebied is. Die definisies was naamlik in natuurlike taal, soos 'n onderwyser 'n woord se betekenis aan 'n aanleerder sou verduidelik. Daarbenewens is die definisies ook in volsinformaat aangebied. Hierdie sogenaamde COBUILD-metode is sedertdien met vrug in talle ander woordeboeke gebruik, onder meer in die Afrikaanse aanleerderwoordeboek *Basiswoordeboek van Afrikaans*, asook in een van die woordeboeke wat in hierdie artikel bespreek is, naamlik Odendal se *Kernwoordeboek*. Hierdie tipe betekenisparafrase maak dit vir die teikengebruiker moontlik om betekenisinligting op 'n maklike manier aan die aangebode data te onttrek. Daar is geen sintaktiese hinderlikhede in die definisies nie en die didaktiese aard van die definisies verseker duidelike begrip van, maar ook insae in natuurlike sinsvorming in Afrikaans. Die *Kernwoordeboek* volg hier 'n werkswyse wat ongelukkig nie genoegsaam in die ander woordeboeke toegepas word nie. Ook met die keuse van hierdie definisietipe moet leksikograwe die taalgebruik in en sinskonstruksie van die betekenisparafrase aanpas by die vaardighede en taalvermoë van die woordeboek se geïdentifiseerde teikengebruikers.

Die gebruik van volsindefiniesies in natuurlike taal het dikwels ruimte-

implikasies aangesien so 'n definisie dikwels meer plek nodig het as 'n sinsnede of woordgroep in tipiese leksikografees. Ook van die ander aanbevelings wat in die loop van hierdie artikel gemaak is, byvoorbeeld dat die verklaring van elke betekenisonderskeiding van 'n polisemiese woord op 'n nuwe reël begin, het ruimte-implikasies. Die skrywers is bewus van hierdie werklikheid van die leksikografiese praktyk maar in die aanbieding van voorstelle in die teoretiese leksikografie moet daar in die eerste plek gewerk word met wat die beste vir die woordeboekgebruiker sal wees. Alles wat voorgestel word, is nie noodwendig haalbaar in een woordeboek nie maar uitgewers en praktiserende leksikograwe kan wel deeglik kennis neem van die voorstelle vanuit die teoretiese leksikografie. Leksikograwe moet egter daarteen waak om deur gebrek aan ruimte verplig te word om 'n te kompakte data-aanbod te gebruik wat weens 'n te hoë mate van tekscondensasie of die gebrek aan 'n behoorlike mikroargitektuur die gebruiker sal benadeel in sy/haar woordeboekgebruik. Die inskrywings in skoolwoordeboeke mag nie die produk wees van 'n reduksieproses wat op inskrywings in omvattender woordeboeke toegepas is nie.

6. Ten slotte

Die ontwikkeling van die teoretiese leksikografie het 'n besliste invloed op die leksikografiepraktyk — ook in Suid-Afrika en ook op Afrikaanse woordeboeke. Die ouer skoolwoordeboeke is dikwels saamgestel in 'n era toe die teoretiese leksikografie nog in 'n vroeë stadium van ontwikkeling was. Hulle is dus nie in dieselfde mate deur teoretiese beginsels gelei as wat vandag die geval is nie. Desondanks vorm hulle 'n noodsaaklike basis vir die verdere ontwikkeling van die leksikografiepraktyk. Danksy die wisselwerking tussen teorie en praktyk toon die meer onlangse skoolwoordeboeke vir huistaalleerders 'n groot verbetering vergeleke met hulle voorgangers. Die gebruikersgerigtheid van hierdie woordeboeke en die erkenning van die betrokke leksikografiese funksies verhoog hulle gehalte as praktiese instrumente. 'n Groter sensitiwiteit vir die diverse aard van die gebruikssituasie en die gebruikerssituasie behoort verdere tipologiese verfyning in die kategorie skoolwoordeboeke mee te bring en 'n meer gerigte lemmakeuse en bewerking sal die teikengebruikers se behoeftes in 'n nog hoër mate bevredig. Die begrip van teoretiese leksikograwe vir die probleme van die praktyk en die bevestiging van praktiserende leksikograwe van die belang van teoreties gefundeerde woordeboeke kan 'n aktiewe wisselwerking verseker wat die gehalte van skoolwoordeboeke aansienlik sal verbeter.

Bibliografie

Woordeboeke

Bergenholtz, Henning et al. (Reds.). 1997. *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.

- Bosman, D.B. en I.W. van der Merwe (Reds.).** 1984⁸. *Tweetalige woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg.
- Bosman, D.B. en I.W. van der Merwe (Reds.).** 2004¹⁰. *Tweetalige skoolwoordeboek. Afrikaans–Engels/Engels–Afrikaans*. Kaapstad: Nasou.
- Combrink, Johan et al. (Reds.).** 1985². *Junior verklarende woordeboek*. Johannesburg: Educum.
- Conradie, Ben.** 1967. *Verklarende Afrikaanse woordeboek vir laerskole*. Kaapstad: Juta.
- Conradie, C. Jac.** 2001. *Shuters Junior woordeboek*. Pietermaritzburg: Shuter en Shooter.
- De Sousa, José Martínez.** 1995. *Diccionario de lexicografía práctica*. Barcelona: Bibliograf.
- Gouws, Ronél et al. (Reds.).** 2010. *Grondslagfasewoordeboek Afrikaans/English*. Kaapstad: Maskew Miller Longman.
- Gouws, Rufus et al. (Reds.).** 2004. *Nuwe woordeboek sonder grense*. Kaapstad: Maskew Miller.
- Gouws, Rufus, et al. (Reds.).** 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- Hartmann, R.R.K. en Gregory James.** 1998. *Dictionary of Lexicography*. Londen/New York: Routledge.
- Kritzinger, M.S.B. et al. (Reds.).** 1921. *Woordeboek: Afrikaans–Engels. Engels–Afrikaans*. Pretoria: J.L. van Schaik.
- Kromhout, Jan.** 2007³⁵. *Skoolwoordeboek. School Dictionary. Afrikaans–Engels. English–Afrikaans*. Kaapstad: Pharos.
- Louw, Phillip et al. (Reds.).** 2007. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. Kaapstad: Oxford University Press.
- Luther, Jana et al. (Reds.).** 2009. *HAT Afrikaanse skoolwoordeboek*. Kaapstad: Pearson Education.
- Odendal, François F.** 2000. *Kernwoordeboek*. Midrand: Perskor.
- Sinclair, John et al. (Reds.).** 1987. *Collins COBUILD English Language Dictionary*. Londen: Collins.

Ander bronne

- Bergenholtz, Henning, Sven Tarp en Herbert E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L., H. Kalverkämper en H.E. Wiegand (Reds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832*. Berlyn/New York: Walter de Gruyter.
- Geeraerts, Dirk.** 1984. Dictionary Classification and the Foundations of Lexicography. *ITL Review* 63: 37-63.
- Gouws, Rufus H.** 1989. *Leksikografie*. Pretoria/Kaapstad: Academica.
- Gouws, Rufus H.** 2007. Op pad na 'n nuwe woordeboektipologie. *Southern African Linguistics and Applied Language Studies* 25(3): 319-331.
- Hausmann, Franz Josef.** 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand en Ladislav Zgusta (Reds.). 1989-1991: 1-19.
- Hausmann, Franz Josef.** 1990. Das Kinderwörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand en Ladislav Zgusta (Reds.). 1989-1991: 1365-1368.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand en Ladislav Zgusta (Reds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*.

- Handbücher zur Sprach- und Kommunikationswissenschaft 5.1–5.3. Berlyn/New York: Walter de Gruyter.
- Hernández, Humberto Hernández.** 1989. *Los diccionarios de orientación escolar. Contribución al estudio de la lexicografía monolingüe española*. Tübingen: Max Niemeyer.
- Kromann, H.-P., T. Riiber en P. Rosbach.** 1984. Überlegungen zur Grundfragen der zweisprachigen Lexikographie. Wiegand, H.E. (Red.). 1984. *Studien zur neuhochdeutschen Lexikographie*: 159-238. Hildesheim: Georg Olms.
- Lagane, René.** 1990. Les dictionnaires scolaires: enseignement de la langue maternelle. Hausmann, Franz Josef, Oskar Reichmann, Herbert E. Wiegand en Ladislav Zgusta (Reds.). 1989–1991: 1368-1378.
- Lombard, Frederik J.** 1990. *'n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke*. Ongepubliseerde D.Litt-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Mabika Mbokou, Ludwine.** 2006. *A Model for the Macro- and Microstructure of a Yipunu–French School Dictionary*. Ongepubliseerde D.Litt-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Malkiel, Y.** 1967. A Typological Classification of Dictionaries on the Basis of Distinctive Features. Householder, F.W. en S. Saporta (Reds.). 1967. *Problems in Lexicography. Report of the Conference on Lexicography held at Indiana University, November 11–12, 1960*: 3-24. Bloomington: Universiteit van Indiana.
- Nyangone Assam, Blanche.** 2006. *Dictionaries as Teaching Instruments for Mother-tongue Education: The Case of Fang in Gabon*. Ongepubliseerde D.Litt-proefskrif. Stellenbosch: Universiteit van Stellenbosch.
- Prinsloo, Daan J. en Gilles-Maurice de Schryver.** Op koms. Do Dictionaries Define on the Level of their Target Users? A Case Study for Three Dutch Dictionaries. *International Journal of Lexicography*.
- Tarp, Sven.** 2000. Theoretical Challenges to LSP Lexicography. *Lexikos* 10: 189-208.
- Tarp, Sven.** 2008. The Third Leg of Two-legged Lexicography. *Hermes. Journal of Linguistics* 40: 117-131.
- Tarp, Sven.** 2009. Homonymy and Polysemy in a Lexicographic Perspective. Bergenholtz, Henning en Thomas Herbst (Reds.). 2009. *Classifications in Linguistics and Lexicography. Special Issue of Zeitschrift für Anglistik und Amerikanistik* 57(3): 289-306.
- Tarp, Sven.** 2010. H.A. Welker and Pedagogical Lexicography. *Lexikos* 20: 730-749.
- Tarp, Sven en Rufus H. Gouws.** 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke. *Tydskrif vir Geesteswetenskappe* 44(4): 276-298.
- Tarp, Sven en Rufus H. Gouws.** 2008. A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries. *Lexikos* 18: 232-255.
- Welker, Herbert Andreas.** 2008. *Panorama geral da lexicografia pedagógica*. Brasília: Thesaurus Editora.
- Zgusta, Ladislav.** 1971. *Manual of Lexicography*. Den Haag: Mouton.

Hybride textuelle Strukturen und hybride textuelle Einheiten. Ein Beitrag zur Theorie der Wörterbuchform

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg,
Heidelberg, Bundesrepublik Deutschland*
(*herbert.ernst.wiegand@gs.uni-heidelberg.de*)

Zusammenfassung: In diesem Beitrag wird die Bildung, Darstellung und Leistung von hybriden textuellen Strukturen, die akzessive Einträge aufweisen, am Beispiel von Wörterbuchartikeln behandelt sowie erklärt, welche Eigenschaften hybride textuelle Einheiten haben. Ein Wörterbuchartikel eines Printwörterbuchs weist immer dann neben einer hierarchischen reinen eine hierarchische hybride Textkonstituentenstruktur auf, wenn in ihm mindestens ein funktionaler Angabezusatz auftritt, z.B. ein oben oder unten erweiternder oder ein binnenerweiternder. Da funktionale Angabezusätze Textsegmente mit Angabefunktion aber ohne Textkonstituentenstatus sind, werden sie durch nichtfunktionale Segmentation ermittelt, so dass neben funktionalen auch nichtfunktionale Textsegmente gegeben sind, die dann bei der Strukturbildung in die Trägermengen eingehen, so dass die Trägermengen aller hybriden hierarchischen Strukturen elementenheterogen, während die Trägermengen aller hierarchischen reinen Strukturen elementenhomogen sind. In den Strukturgraphen für hierarchische hybride Artikelstrukturen sind dann die Knoten für diejenigen Textsegmente, die den Hybridstatus der Strukturen bewirken, entweder durch Pfeilkanten für die textarchitektonischen *oberhalb/unterhalb*-Relationen mit den Knoten für die Textkonstituenten verbunden, so dass die Strukturgraphen architektonisch angereichert sind, oder durch besonders markierte Kanten, die die Knoten für die nichtfunktionalen Textsegmente und die für die binnenerweiternden funktionalen Angabezusätze mit den Knoten für die Textkonstituenten verbinden. Zu jedem Typ von hierarchischer reiner Artikelstruktur gehören mehrere Typen von hybriden Artikelstrukturen; entsprechendes gilt für hierarchische reine Angabestrukturen. Nur eine Auswahl aus den Typologien der hybriden Artikel- und Angabestrukturen wird behandelt sowie eine kleine Auswahl von hybriden textuellen Einheiten, die kriteriale Eigenschaften von zwei Textsegmentklassen aufweisen (vgl. die Stichwörter).

Stichwörter: ANGABERELATION, ELEMENTENHETEROGENE TRÄGERMENGE, FUNKTIONALER ANGABEZUSATZ, FUNKTIONAL-POSITIONALE SEGMENTATION, HIERARCHISCHE ARCHITEKTONISCH ANGEREICHERTE ARTIKELMIKROSTRUKTUR, HIERARCHISCHE HYBRIDE ANGABENKONSTITUENTENSTRUKTUR MIT GLOSSATBEDINGTER TEILSTRUKTUR, HIERARCHISCHE HYBRIDE ANGABENSTRUKTUR, HIERARCHISCHE HYBRIDE ARTIKELKONSTITUENTENSTRUKTUR, HIERARCHISCHE HYBRIDE ARTIKELMIKROSTRUKTUR, HIERARCHISCHE HYBRIDE EXHAUSTIVE ANGABENSTRUKTUR, HIERARCHISCHE HYBRIDE GLOSSATBEDINGTE ANGABESTRUKTUR, HIERARCHISCHE HYBRIDE FLACHE

DOPPELGLOSSATBEDINGTE ANGABESTRUKTUR, HIERARCHISCHE HYBRIDE MINIMIERTE GLOSSATBEDINGTE ANGABESTRUKTUR, HIERARCHISCHE HYBRIDE TEXTKONSTITUENTENSTRUKTUR, HIERARCHISCHE HYBRIDE TIEFE DOPPELGLOSSATBEDINGTE ANGABESTRUKTUR, HIERARCHISCHE REINE TEXTKONSTITUENTENSTRUKTUR, HYBRIDE VERWEISKENNZEICHNUNG, NICHTFUNKTIONALE-POSITIONALE SEGMENTATION, ORDUNGSRELATION, SEGMENTATIVE ISOLIERUNG, VERTIKALE ANGABEARCHITEKTUR

Abstract: Hybrid textual structures and hybrid textual units. A contribution to the theory of dictionary structures.

In this contribution, the formation, presentation and performance of hybrid textual structures that display accessible entries are discussed by using examples from dictionary articles. The features of hybrid textual units are also explained. A dictionary article in a printed dictionary always displays both a hierarchical pure and a hierarchical hybrid text constituent structure, when it contains at least one functional item addition, e.g. an upward- or downward- or an internally-expanded one. Because functional item additions are text segments with an item function but without text constituent status, they are enabled by means of non-functional segmentation, so that both functional and non-functional text segments prevail. During the formation of structures they then enter the structure-carrying set so that the structure-carrying set of all hybrid hierarchical structures are element-heterogeneous whilst the structure-carrying set of all hierarchical pure structures are element-homogeneous. In the structural diagrams of hierarchical hybrid article structures, the nodes for those text segments that establish the hybrid status of the structures are connected with the nodes for the text constituents either by means of arrows for the text-architectonic *upward/downward* relations, so that the structural graphs are architectonically enriched, or by means of specially marked edges that connect the nodes for the non-functional text segments and those for the internally-expanded functional item additions with the nodes for the text constituents. To each type of hierarchical pure article structure belong various types of hybrid article structures. The same applies to hierarchical pure item structures. Only a selection from the typology of hybrid article and item structures are discussed as well as a small selection of hybrid textual units that display determining features of two text segment classes (cf. the keywords).

Keywords: ELEMENT-HETEROGENEOUS STRUCTURE-CARRYING SET, FUNCTIONAL ITEM ADDITION, FUNCTIONAL-POSITIONAL SEGMENTATION, HIERARCHICAL ARCHITECTONICALLY ENRICHED ARTICLE MICROSTRUCTURE, HIERARCHICAL HYBRID ARTICLE MICROSTRUCTURE, HIERARCHICAL HYBRID DEEP DOUBLE GLOSS-CONDITIONED ITEM STRUCTURE, HIERARCHICAL HYBRID EXHAUSTIVE ITEMS STRUCTURE, HIERARCHICAL HYBRID GLOSS-CONDITIONED ITEM STRUCTURE, HIERARCHICAL HYBRID ITEMS CONSTITUENT STRUCTURE WITH GLOSS-CONDITIONED PARTIAL STRUCTURE, HIERARCHICAL HYBRID ITEMS STRUCTURE, HIERARCHICAL HYBRID MINIMISED GLOSS-CONDITIONED ITEM STRUCTURE, HIERARCHICAL HYBRID SHALLOW DOUBLE GLOSS-CONDITIONED ITEM STRUCTURE, HIERARCHICAL HYBRIDARTICLE CONSTITUENT STRUCTURE, HIERARCHICAL HYBRIDTEXT CONSTITUENT STRUCTURE, HIERARCHICAL PURE TEXT CONSTITUENT STRUCTURE, HYBRID CROSS-REFERENCE MARKER, ITEM RELATION, NON-FUNCTIONAL-POSITIONAL SEGMENTATION, ORDERING RELATION, SEGMENTATIVE ISOLATION, VERTICAL ITEM ARCHITECTURE

1. Vorbemerkung: Was meint das Prädikat *hybrid*?

Dt. *Hybride* ist nach lat. *hybrida* 'Mischling' gebildet; das zugehörige deutsche Adjektiv *hybrid* weist ein Bedeutungsspektrum auf, das nach Duden-⁴GFWB wie folgt paraphrasiert werden kann: „gemischt, von zweierlei Herkunft, aus Verschiedenem zusammengesetzt; durch Kreuzung, Mischung entstanden“. Besonders zutreffend ist hier zunächst die Bedeutungsparaphrasenangabe *von zweierlei Herkunft*, und zwar aus folgendem Grund: Die Herkunft der Einheiten der Theorie der Wörterbuchform ist immer das Ergebnis einer korrekten Methodenanwendung. Sind die textuellen Einheiten, mit denen die Theorie der Wörterbuchform arbeitet und die im Zuge der Strukturbildung zu einer Trägermenge zusammengefasst werden, das Ergebnis der Anwendung mindestens zweier verschiedener metalexikographischer Methoden, dann gelten die textuellen Strukturen als *hybrid*. Sind die Elemente der Trägermenge dagegen das Ergebnis der Anwendung nur einer Methode, dann gelten die Strukturen als *rein* (oder *nichthybrid*). Nicht nur textuelle Strukturen, sondern auch textuelle Einheiten können hybrid sein. Dies ist stets dann der Fall, wenn eine Einheit eine Eigenschaft aufweist, die zu den kriterialen Eigenschaften der textuellen Einheiten einer anderen Klasse gehört, aber wegen ihrer restlichen Eigenschaften dieser anderen Klassen nicht zugeordnet werden kann (vgl. Wiegand 2010).

2. Zum Zusammenwirken von Methoden, Theorieausschnitten und Terminologie

Zur Theorie der Wörterbuchform gehört eine Heuristik (vgl. Wiegand 2010), mit der u.a. auch mehrere Segmentationsmethoden für die Segmentation kondensierter lexikographischer Texte bereitgestellt werden (vgl. z.B. Wiegand 1990: 20-26; 2005: 217-226; Wiegand und Fuentes Morán 2009). Für die Anwendung der Segmentationsmethoden sind Anwendungskonventionen und Korrektheitsbedingungen festgelegt. Für die wissenschaftstheoretisch notwendige Verknüpfung von Methoden und Theorieausschnitten gilt neben anderen das folgende Prinzip: Als Elemente für konkrete textuelle Wörterbuchstrukturen kommen ausschließlich solche Textsegmente in Betracht, die durch die korrekte Anwendung einer der Segmentationsmethoden erhältlich sind. Es gelten folgende Unterscheidungen:

- (i) Durch eine Anwendung der Methode der exhaustiven funktional-positionalen Segmentation, die eine der Varianten der Methode der funktional-positionalen Segmentation darstellt, sind alle Arten von Textkonstituenten erhältlich, also nur solche funktionalen Textsegmente, die eine diskrete kontinuierliche Textsegmentform, mindestens eine genuine Textsegmentfunktion und genau eine feste textuelle Position innerhalb (und nicht oberhalb oder unterhalb) der sprachlichen Kette des Textes

aufweisen; dies sind die Folgenden: Angaben, Angabetexte und nichttypographische Strukturanzeiger (vgl. z.B. Wiegand 2009: 254ff; 2009a: 65ff).

- (ii) Durch eine Anwendung der Methode der nichtfunktional-positionalen Segmentation sind alle nichtfunktionalen Textsegmente sowie alle binernerweiternden funktionalen Angabezusätze erhältlich (vgl. z.B. Wiegand 2006: 201ff; 2009: 268ff).
- (iii) Durch eine Anwendung der Methode der funktionalen segmentativen Isolierung sind alle oben und unten erweiternden funktionalen Angabezusätze, also die ohne eigene Position in der sprachlichen Kette, erhältlich (vgl. z.B. Wiegand 2006: 193ff; 2009: 270ff).

Es gelten weiterhin die folgenden Festlegungen:

- (a) Konkrete hierarchische Textkonstituentenstrukturen, deren Trägermengen in dem Sinne elementenhomogen sind, dass sie als Elemente nur Textsegmente mit Textkonstituentenstatus aufweisen (vgl. (i)) heißen *konkrete reine* (oder: *nichthybride*) *Textkonstituentenstrukturen*: Alle ihre Elemente sind in dem Sinne von gleicher Herkunft, dass sie durch Anwendungen der gleichen metalexikographischen Methode erhältlich sind. Entsprechend heißen abstrakte hierarchische Textkonstituentenstrukturen, deren Trägermengen nur Klassen von Textsegmenten mit Textkonstituentenstatus aufweisen, *abstrakte reine* (oder: *nichthybride*) *Textkonstituentenstrukturen*.
- (b) Konkrete hierarchische Textkonstituentenstrukturen, deren Trägermengen in dem Sinne elementenheterogen sind, dass sie als Elemente sowohl Textsegmente mit als auch solche ohne Textkonstituentenstatus aufweisen (vgl. (i) bis (iii)), heißen *konkrete hybride Textkonstituentenstrukturen*: Ihre Elemente sind in dem Sinne von verschiedener Herkunft (vgl. die in 1. zitierte Paraphrasierung im Duden-⁴GFwB), dass sie durch Anwendungen verschiedener metalexikographischer Methoden erhältlich sind. Entsprechend heißen abstrakte hierarchische Textkonstituentenstrukturen, deren Trägermengen sowohl Klassen von Textsegmenten mit Textkonstituentenstatus als auch Klassen von Textsegmenten ohne Textkonstituentenstatus aufweisen, *abstrakte hybride Textkonstituentenstrukturen*. — Es kann daher festgestellt werden: Kriterium für die Unterscheidung von konkreten hierarchischen reinen und konkreten hierarchischen hybriden Textkonstituentenstrukturen ist die Zusammensetzung der Trägermengen: Konkrete hierarchische reine Textkonstituentenstrukturen weisen elementenhomogene, konkrete hierarchische hybride Textkonstituentenstrukturen weisen elementenheterogene Trägermengen auf. Für die abstrakten Strukturen gelten diese Zuordnungen *mutatis mutandis*. Ergänzt sei: Auch hierarchische textuelle Strukturen, die keine Text-

konstituentenstrukturen sind, wie z.B. bestimmte Angabestrukturen, werden mittels des gleichen Kriteriums in reine und hybride Strukturen unterteilt (vgl. 5).

3. Reine artikelinterne Textkonstituentenstrukturen

In kondensierten Wörterbuchartikeln (*sensu* Wiegand 2003: 203ff) treten reine Textkonstituentenstrukturen auf, die zu fünf Typen gehören: zum Typ der hierarchischen reinen Artikelkonstituentenstruktur, zum Typ der hierarchischen reinen Artikel mikrostruktur, zum Typ der hierarchischen reinen exhaustiven Angabenstruktur, zum Typ der hierarchischen reinen Angabenstruktur und zum Typ der hierarchischen artikelinternen Suchbereichsstruktur (*sensu* Wiegand 2000: 269ff). Letztere werden in diesem Beitrag nicht berücksichtigt. In diesem Abschnitt wird zunächst ein Beispiel für hierarchische reine Artikel mikrostrukturen in etwas vereinfachter Weise behandelt; auf die Grundlagen der Theorie kann dabei allerdings nicht eingegangen werden (vgl. dazu Wiegand 1989 u. 1989a). Die Darstellung in diesem Abschnitt dient lediglich dem Zweck, den grundlegenden Unterschied zwischen reinen und hybriden textuellen Strukturen genauer zu verstehen; sie dient außerdem dazu, auch anhand eines Beispiels eine größere Anschaulichkeit zu erreichen. Gegeben seien die folgenden Wörterbuchartikel:

- wa₁: **Smoking**, der; -s, -s *meist schwarzes Herrenjackett mit seidenen Rockauschlägen für festliche Anlässe: im S. erscheinen*
- wa₂: **Streber**, der; -s, - jmd., *der bes. in der Ausbildung ehrgeizig, aber egoistisch auf sein Fortkommen hinarbeitet: ein gewissenloser, widerlicher S.*
- wa₃: **Skale**, die; -, -n *(bezahlte) Maßeinteilung an Meßgeräten: einen Meßwert an einer S. ablesen*
- wa₄: **Sonate**, die; -, -n *Komposition mit meist drei od. vier Sätzen für ein od. mehrere Soloinstrumente: sie spielten eine S. für Violine und Klavier von Beethoven*
- wa₅: **Spezialist**, der; -en, -en *jmd., der über besondere Kenntnisse, Fähigkeiten auf einem Fachgebiet verfügt, Fachmann: er ist S. für Halbleitertechnik; er muß einen Spezialisten (Facharzt) aufsuchen*
- wa₆: **Sweiß**, der; -es. /o. Pl./ **1.** *bes. bei Hitze aus den Poren der Haut austretende wäßrige Absonderung bestimmter Drüsen: der S. steht ihm auf der Stirn, brach ihm aus allen Poren; in S. kommen, geraten; ◇ das hat viel S. (Mühe, Anstrengung) gekostet – 2. Jägerspr. Blut des Wildes und des Hundes*

+ im Schweiß seines Angesichts (*mit großer Mühe, Anstrengung*)

wa₇: **Bub**, der; -en, en süddt. österr. schweiz.
Knabe, Junge: ein kleiner, wilder B.; ein B. von
fünf Jahren

Abb. 1: Wörterbuchartikel wa₁ bis wa₇ aus HWDG

Im Folgenden betrachten wir zunächst die konkrete hierarchische reine Artikelmikrostruktur von wa₁. Zur Trägermenge jeder Art von konkreter hierarchischer Artikelmikrostruktur gehören keine nichttypographischen Mikrostrukturanzeiger; vielmehr gehören Letztere zur Trägermenge von konkreten hierarchischen Artikelkonstituentenstrukturen, deren prominenteste Substrukturen die konkreten hierarchischen Artikelmikrostrukturen sind sowie zur Trägermenge von konkreten hierarchischen exhaustiven Angabenstrukturen. Demgemäß muss wa₁, damit seine konkrete hierarchische Artikelmikrostruktur erhältlich ist, in einem *ersten methodischen Schritt* so segmentiert werden, dass alle Textkonstituenten gegeben sind außer den nichttypographischen Mikrostrukturanzeigern. Da wa₁ keinen Angabetext aufweist, sind dies nur elementare und nichtelementare Angaben. Daher wird auf wa₁ die Methode der nichtexhaustiven funktional-positionalen Segmentation angewandt, die eine zweite Variante der Methode der funktional-positionalen Segmentation darstellt und zu deren Korrektheitsbedingungen es gehört, dass als Segmentationsergebnisse nur die Angaben gelten. Es gilt die Segmentationskonvention K₁, die besagt, dass Divise (wie z.B. die in den beiden Angaben „-s“ in wa₁) als Teile der Angabeform von Angaben mit Kohäsionsanweisung gelten. Eine andere Segmentationskonvention K₂, die hier nicht in Geltung gesetzt wird, könnte lauten, dass die beiden Divise als unmittelbare Teilangaben, nämlich als verdichtete Wortstammangaben gelten sollen, so dass die beiden „-s“ in wa₁ keine elementaren, sondern nichtelementare Angaben wären.

Im Folgenden wird das Segmentationsprocedere nicht im Detail beschrieben (vgl. dazu z.B. Wiegand 2000: 235ff u. 2009: 256ff), sondern nur das Segmentationsergebnis so aufgelistet, dass jeder elementaren und jeder nichtelementaren Angabe, die wa₁ aufweist, zugleich ein Kleinbuchstabe, der in runden Klammern voran steht, als Angabename zugeordnet wird, wobei zu beachten ist, dass Angabennamen Individuennamen sind; weiterhin wird jede Angabe der Klasse von Angaben mit gleicher allgemeiner wörterbuchgegenstandsbezogener genuiner Angabefunktion zugeordnet, zu der sie gehört. Letzteres geschieht mittels Aussagen der Form „ $a \in \beta$ “, mit „ α “ als Variable für Angabennamen und „ β “ als Variable für Klassensymbole, die Abkürzungen von Angabeklassennamen sind; dabei ist „ \in “ das Symbol für die Element-Klassen-Relation in der Bedeutung *ist ein Element von* oder in der Bedeutung *ist enthalten in*.

- (a) **Smoking**; der; -s, -s [_iAB_j]; $a \in FK$ (= Formkommentar); AB = Angabeblock.

- (b) **Smoking**; $b \in \text{LZGA} \mid \text{Wak} \mid \text{VQK.L} \mid \text{WFA.NSg} \mid \text{RA}$ (= Lemmazeichengestaltangabe, ausgeprägt als lemmatische Substantivangabe, unten erweitert durch eine Wortakzentkennzeichnung (Wak), die zugleich („|“) eine Vokalquantitätskennzeichnung zur Länge (VQK.L) ist, weiterhin zugleich Wortformangabe für den Nominativ Singular (WFA.NSg) und zugleich Rechtschreibangabe (RA)).
- (c) der; -s, -s; $c \in \text{MorA.S}$ (= Morphologieangabe bei Substantiven).
- (d) der; $d \in \text{ArtA} \dashv \text{G} \mid \text{WAr}$ (= Artikelangabe, anhand derer das Genus (= G) und zugleich die Wortart (= WAr) erschließbar ist).
- (e) -s, -s; $e \in \text{DekKA}$ (= Deklinationsklassenangabe).
- (f) -s; $f \in \text{v.SgbA}$ (= verdichtete Singularbildungsangabe).
- (g) -s; $g \in \text{v.PlbA}$ (= verdichtete Pluralbildungsangabe).
- (h) [_iAB_j]; $h \in \text{A-rAus}$ (= Angabe zur regelmäßigen Aussprache; diese ist eine Nullangabe, die in der konkreten Struktur durch einen Angabeblank (_iAB_j) repräsentiert wird, der durch die Belegung seiner beiden Nachbarschaftsvariablen „i“ und „j“ auch positional identifiziert ist).
- (i) *meist schwarzes Herrenjackett mit seidenen Rockaufschlägen für festliche Anlässe*; im S. erscheinen; $i \in \text{SK}$ (semantischer Kommentar).
- (j) [_iAB_j]; $j \in \text{MonA}$ (= Monosemieangabe; diese ist auch eine Nullangabe, die in der konkreten Struktur durch einen Angabeblank (_iAB_j) repräsentiert wird, der durch die Belegung seiner beiden Nachbarschaftsvariablen „i“ und „j“ positional identifiziert ist).
- (k) *Meist schwarzes Herrenjackett mit seidenen Rockaufschlägen für festliche Anlässe*; im S. erscheinen; $k \in \text{SSK}$ (= semantischer Subkommentar).
- (l) [_iAB_j]; *meist schwarzes Herrenjackett [...] Anlässe*; $l \in \text{PragsemA}$ (= pragmatisch-semantische Angabe).
- (m) [_iAB_j]; $k \in \text{A-pragNM}$ (= Angabe zur pragmatischen Nullmarkierung; diese ist ebenfalls eine Nullangabe).
- (n) *Meist schwarzes Herrenjackett [...] Anlässe*; $n \in \text{BPA}$ (= Bedeutungsparaphrasenangabe).
- (o) im S. erscheinen; $o \in \text{v.BeiA}$ (= verdichtete Kompetenzbeispielsangabe).

Die Angaben b, d, f, g, h, j, m, n und o sind elementare Angaben, also solche, die nicht weiter funktional-positional segmentierbar sind. Nullangaben gelten als elementare Angaben, die auslagerungsresistent sind. Die Angaben a, c, e, i, k, l sind nichtelementare Angaben; sie sind mithin restfrei funktional-positional segmentierbar.

In einem *zweiten methodischen Schritt* bilden wir nun die elementenhomogene Trägermenge für $MiS_n^k(wa_1)$, die konkrete (k) hierarchische (h) reine (r) Artikelmikrostruktur (MiS) von wa_1 ; es handelt sich bei der Trägermenge um die um den ganzen Wörterbuchartikel wa_1 erweiterte Menge aller methodisch ermittelnden Angaben a bis o. Die Trägermenge heie $M_{TMiS}^k(wa_1)$; sie kann wie folgt mit einer Mchtigkeit von | 16 | angegeben werden:

$$M_{TMiS}^k(wa_1) = \{wa_1, a, b, c, d, e, f, g, h, i, j, k, l, m, n, o\}.$$

Dass wa_1 ein Element von $M_{TMiS}^k(wa_1)$ ist, macht die Trgermenge nicht zu einer elementenheterogenen Menge, denn auch wa_1 ist eine Textkonstituente, die durch eine Anwendung der Methode der nichtexhaustiven funktionalen-positionalen Segmentation auf die Artikelstrecke S erhltlich ist. Auerdem kann wa_1 in einem formalen Sinn als die grte Angabe aufgefasst werden.

In einem *dritten methodischen Schritt* definieren wir auf $M_{TMiS}^k(wa_1)$ eine zweistellige irreflexive (und damit asymmetrische) sowie transitive Relation vom Typ der Przedenzrelation — sie heie $R_p^k(wa_1)$ — mit dem Relationsterm *s geht voraus y*, mit „x“ und „y“ als Variable fr Angaben. $R_p^k(wa_1)$ gehrt zu den irreflexiven partiellen Ordnungsrelationen. Es gilt: $R_p^k(wa_1) \subseteq M_{TMik}^k(wa_1) \times M_{TMik}^k(wa_1)$. Die strukturprgende Relation $R_p^k(wa_1)$ enthlt als Elemente alle diejenigen 2-Tupel, wie z.B. $\langle a, d \rangle$ und $\langle f, g \rangle$, die — wenn deren Koordinaten in den Relationsterm *x geht voraus y* eingesetzt werden — wahre Stze liefern, wie z.B. *a geht voraus d* (\approx **Smoking** geht voraus der; mit „ \approx “, in der Bedeutung von *entspricht*). $R_p^k(wa_1)$ legt damit vollstndig fest, welche Angaben von wa_1 welchen anderen Angaben von wa_1 vorausgehen. Entsprechendes gilt auch von anderen Relationen (bei spteren Beispielen), die zum Typ der Przedenzrelation gehren. Przedenzrelationen heien auch *Vorgnger-Nachfolger-Relationen*.

In einem *vierten methodischen Schritt* definieren wir auf der Trgermenge $M_{TMiS}^k(wa_1)$ eine zweistellige reflexive, antisymmetrische (oder: identitive) und transitive Relation vom Typ der partitiven Relation — sie heie $R_{part}^k(wa_1)$ — mit dem Relationsterm *x ist eine Angabe als Teil von y*, mit „x“ als Variable fr Angaben und „y“ als Variable fr Angaben und wa_1 . Es gelten die Aussagen: $R_{part}^k(wa_1) \subseteq M_{TMik}^k(wa_1) \times M_{TMik}^k(wa_1)$ und $R_p^k(wa_1) \cap R_{part}^k(wa_1) = \emptyset$. Die Relation $R_{part}^k(wa_1)$ gehrt zu den reflexiven partiellen Ordnungsrelationen. Sie enthlt als Elemente diejenigen 2-Tupel (oder: geordneten Paare), wie z.B. $\langle d, c \rangle$ und $\langle a, wa_1 \rangle$, die — wenn deren Koordinaten in den Relationsterm *x ist eine Angabe als Teil von y* eingesetzt werden — wahre Stze, wie z.B. *d ist eine Angabe als Teil von c* (\approx „der“ ist eine Angabe als Teil von „der“ -es, -e“), liefern. Die Relation $R_{part}^k(wa_1)$ legt damit vollstndig fest, (i) welche Angaben von wa_1 Teilangaben von welchen anderen Angaben von wa_1 sind, sowie (ii) welche Angaben Teile von wa_1 sind. Die Vereinigung der beiden strukturprgenden Relationen, $R_p^k(wa_1) \cup R_{part}^k(wa_1)$, ordnet die Trgermenge vollstndig.

Nach der Ausführung des vierten methodischen Schrittes ist $rMiS_h^k(wa_1)$, die reine konkrete hierarchische Artikelmikrostruktur von wa_1 , als eine Ordnungsstruktur gegeben, die festlegt, welchen strukturellen Ort jede Angabe in $rMiS_h^k(wa_1)$ aufweist. Damit kann festgestellt werden, dass hierarchische reine Artikelmikrostrukturen von vollständig kondensierten Wörterbuchartikeln (*sensu* Wiegand 2003: 207f), die keine Angabetexte aufweisen, als artikelinterne Angabedistributionsstrukturen gelten können.

Der *fünfte methodische Schritt* besteht darin, dass die Definition für die konkrete hierarchische reine Artikelmikrostruktur von wa_1 formuliert wird. Dabei muss auf die so genannten Hilfsbasismengen (in der Terminologie von Bourbaki 1957) zurückgegriffen werden. Hilfsbasismengen sind endliche Mengen von Objekten aus dem Bereich, auf den die mathematische Strukturtheorie der Mathematikergruppe Bourbaki angewandt wird. Der Bereich ist hier das HWDG. Für die Analyse von hierarchischen reinen Artikelmikrostrukturen in konkreter und abstrakter Ausprägung benötigt man folgende Hilfsbasismengen:

- WA_1 ist die Menge aller kondensierten Wörterbuchartikel des HWDG; ($wa_1 \in WA_1$)
- A_1 ist die Menge aller Angaben in kondensierten Wörterbuchartikeln des HWDG
- AK_1 ist die Menge aller Mengen von Angaben mit gleicher allgemeiner genuiner Funktion, die in kondensierten Wörterbuchartikeln des HWDG gegeben sind; kurz: AK_1 ist die Menge aller Angabeklassen. Die Elemente von AK_1 sind Äquivalenzklassen über A_1 , die dadurch gebildet werden, dass auf A_1 eine Relation vom Typ der Äquivalenzrelation definiert wird, die transitiv, symmetrisch und reflexiv ist und zu der der Relationsterm *x ist mit y gleich hinsichtlich der allgemeinen wörterbuchgegenstandsbezogenen genuinen Funktion* gehört, mit „x“ und „y“ als Variablen für Angaben aus A. Es gilt: $AK_1 \subseteq A_1$.

Die Definition lautet dann wie folgt:

Def. 1: $rMiS_h^k(wa_1)$, die konkrete (k) hierarchische (h) reine (r) Artikelmikrostruktur (MiS) von wa_1 ist eine Ordnungsstruktur, bestehend aus der Trägermenge $M_{TMiS}^k(wa_1)$, auf der die partitive Relation $R_{part}^k(wa_1)$ sowie die Präzedenzrelation $R_p^k(wa_1)$ definiert sind, und für die folgende Beziehungen zu den Hilfsbasismengen WA_1 , A_1 und AK_1 gelten:

- (i) Ein Element stammt aus WA_1 .
- (ii) Die anderen Elemente stammen aus Elementen von AK_1 (also aus den Äquivalenzklassen über A_1).

Eine formale Form der Def. 1 kann wie folgt angegeben werden:

Def. 1' : $\text{rMiS}_h^k(wa_1) = \text{def. } \langle M_{\text{TMIS}}^k(wa_1), A_1, AK_1, WA_1; R_{\text{part}}^k(wa_1), R_p^k(wa_1) \rangle$.

Bei einer strengen Anwendung der mathematischen Strukturtheorie von Bourbaki gehört zu einer Definition von Ordnungsstrukturen eine Menge von Ordnungsaxiomen; diese sind Bedingungen, die alle konkreten hierarchischen reinen Artikelmikrostrukturen erfüllen müssen, wenn für sie die Def. 1 gelten soll. Informell formuliert lauten solche Bedingungen wie folgt:

- (a) Der Formkommentar von wa_1 ist eine unmittelbare Textkonstituente von wa_1 und hat keinen Vorgänger bezüglich $R_p^k(wa_1)$.
- (b) Der semantische Kommentar von wa_1 ist eine unmittelbare Textkonstituente von wa_1 und hat genau einen Vorgänger bezüglich $R_p^k(wa_1)$, nämlich den Formkommentar.
- (c) Das aus WA stammende Element wa_1 dominiert alle anderen Elemente der Trägermenge.

Die relativ aufwendige formale Formulierung aller Ordnungsaxiome kann man sich sparen, wenn man den Strukturgraphen für $\text{rMiS}_h^k(wa_1)$ (am besten zusammen mit dem für die zugehörige abstrakte (a) Artikelmikrostruktur $\text{rMiS}_h^a(wa_1)$) angibt. Jeder, der weiß, wie man einen solchen geordneten Strukturgraphen, der aus zwei aufeinander abgebildeten geordneten Baumgraphen besteht, richtig liest, kann leicht alle Ordnungsaxiome an diesem ablesen.

Abstrakte Artikelmikrostrukturen sind dadurch erhältlich, dass die beiden strukturprägenden Ordnungsrelationen auf einer Trägermenge definiert werden, die nicht wa_1 und die konkreten Angaben a bis o enthält, sondern die Klasse WA_1 der Wörterbuchartikel des HWDG sowie die Angabeklassen, zu denen die Angaben a bis o gehören. Die Trägermenge — sie heiße $M_{\text{TMIS}}^a(wa_1)$ — kann unter Verwendung der in (a)–(o) eingeführten Klassensymbole entsprechend wie folgt mit einer Mächtigkeit von $|16|$ angegeben werden:

$$M_{\text{TMIS}}^a(wa_1) = \{WA, FK, LZGA \sqcup WAK \mid VQK.L \mid WFA.NSg \mid RA, MorA.S, ArtA \rightarrow G \mid WAr, DekKA, v.SgbA, v.PLbA, A-rAus, SK, MonA, SSK, PragsemA, A-pragNM, v.BPA, v.KBeiA\}.$$

Der einfach kommentierte Strukturgraph zur konkreten (und zur isomorphen abstrakten) hierarchischen reinen Artikelmikrostruktur von wa_1 findet sich in Abb. 2.

Es lassen sich zahlreiche Typen von reinen Mikrostrukturen unterscheiden. Eine (nicht ganz vollständige) Übersicht findet man in Wiegand (2002: 573-580).

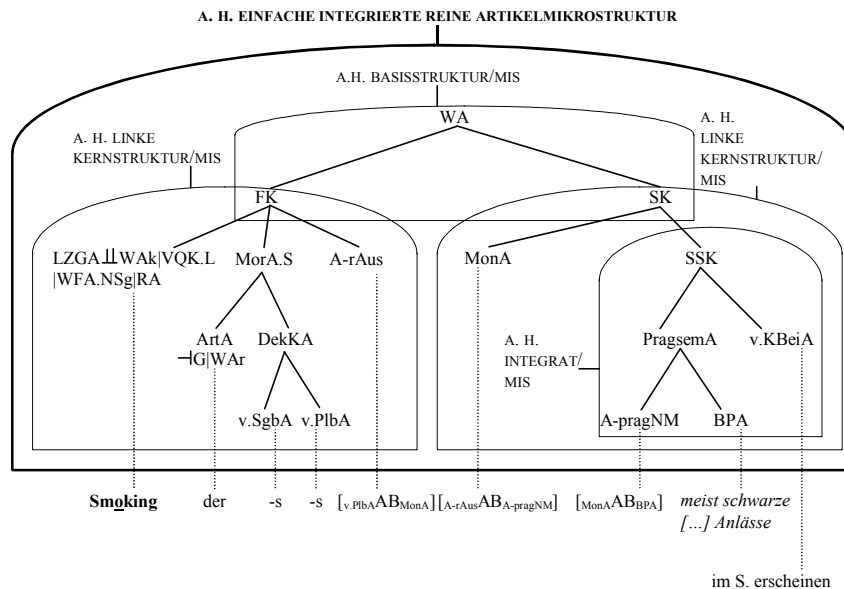


Abb. 2: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen einfachen integrierten reinen Artikelmikrostruktur, die wa₁ in Abb. 1 aufweist.

Darstellungskonventionen: „x — y“ bedeutet (von unten nach oben gelesen) soviel wie x ist eine Angabe als Teil von y; „u - - V“ bedeutet (von unten nach oben gelesen) soviel wie u ist ein Element von V (= u ∈ V, wobei „V“ ein Klassensymbol ist. „|“ bedeutet soviel wie zugleich; LINKE KERNSTRUKTUR/MIS ist zu lesen wie linke Kernstruktur als Teilstruktur der Artikelmikrostruktur; Abkürzungen (alle Abkürzungen außer WA, mit denen Knoten etikettiert sind, sind Klassensymbole für Klassen von Angaben mit gleicher allgemeiner wörterbuchgegenstandsbezogener genuiner Angabefunktion): A. = ABSTRAKTE; H. = HIERARCHISCHE; WA = Wörterbuchartikel; FK = Formkommentar; SK = semantischer Kommentar; LZGA = Lemmazeichengestaltangabe; WFA.NSg = Wortformenangabe Nominativ Singular; RA = Rechtschreibangabe; MorA.S = Morphologieangabe bei Substantiven; A-rAus = Angabe zur regelmäßigen Aussprache; ArtA = Artikelangabe; DekKA = Deklinationsangabe; v.SgbA = verdichtete Singularbildungsangabe; v.PlbA = verdichtete Pluralbildungsangabe; PragsemA = pragmatisch-semantische Angabe; A-pragNM = Angabe zur pragmatischen Nullmarkierung; BPA = Bedeutungsparaphrasenangabe; v.KBeiA = verdichtete Kompetenzbeispielangabe; iAB_j = Angabeblank (mit Nachbarschaftsvariablen „i“, „j“, deren Belegung die Identifizierung und Positionierung des AB sicherstellen)

4. Hybride artikelinterne Textkonstituentenstrukturen I: Artikelkonstituenten- und Artikelmikrostrukturen

Hybride Artikelkonstituenten- und Artikelmikrostrukturen lassen sich einem kondensierten Wörterbuchartikel (i.S.v. Wiegand 2003: 200f) dann zuweisen,

wenn er mindestens einen funktionalen Angabezusatz aufweist. Für Wörterbuchartikel, in denen funktionale Angabezusätze gegeben sind, lässt sich sowohl eine reine als auch mindestens eine hybride Artikelkonstituenten- und Artikelmikrostruktur angeben. Funktionale Angabezusätze sind funktionale Textsegmente, mit denen — wie mit Angaben — etwas angegeben wird, so dass sie eine Angabefunktion haben, die aber — im Unterschied zu Angaben — keinen Textkonstituentenstatus aufweisen (vgl. zu funktionalen Angabezusätzen z.B. Wiegand 2005: 326-330; 2007: 192ff); sie sind nicht — wie Angaben — als Ergebnis einer funktional-positionalen Segmentation erhältlich, weil sie entweder keine eigene Position in der sprachlichen Kette aufweisen (wie z.B. der Unterstrich unter dem Angabeformsegment „o“ in der Lemmazeichengestaltangabe von wa₁), oder weil sie, wenn sie eine eigene Position in der sprachlichen Kette aufweisen, nicht funktional-positional isolierbar sind, wie z.B. das semantisch Binnenglossat „*Facharzt*“ in wa₅ in Abb.1; denn eine Segmentation führt zu den beiden nichtfunktionalen Textsegmenten „er muß einen Spezialisten“ und „aufsuchen“, so dass die Segmentation nichtfunktional-positional ist. Der Unterstrich in wa₁ gehört zu den unten erweiternden funktionalen Angabezusätzen, und zwar handelt es sich um einen unten erweiternden bifunktionalen Angabezusatz, der an das Angabeformsegment „o“ hinaufadressiert ist; mit ihm wird eine Wortakzentkennzeichnung realisiert, die zugleich eine Vokalquantitätskennzeichnung zur Länge ist (Wak | VQK.L). Das Gleiche gilt für die Unterstriche in wa₂ bis wa₄. Weist man wa₅ eine konkrete und eine abstrakte hierarchische reine Artikelmikrostruktur zu, dann sind diese auf die gleiche Weise erhältlich wie die Artikelmikrostrukturen von wa₁ und können wie folgt in Abb. 3 dargestellt werden.

Wenn wir den Strukturgraphen in Abb. 3 aufmerksam betrachten, erkennen wir, dass in der abstrakten hierarchischen Artikelmikrostruktur zu der Klasse, zu der die Wortakzentkennzeichnung, die zugleich eine Vokalquantitätskennzeichnung zur Kürze gehört, kein eigener Knoten gegeben ist. Sie weist damit keinen eigenen strukturellen Ort im geordneten Baumgraphen auf, der mittels mindestens einer Kante mit einem anderen strukturellen Ort des gleichen Baumgraphen verbunden ist. Dass es diesen unten erweiternden bifunktionalen Ansatz in wa₅ gibt, erfährt man nur aus dem Knotenetikett „LZGA□Wak | VQK.K | WFA.NSg | RA“. Auch in der zugehörigen isomorphen konkreten Artikelmikrostruktur kann daher der Unterpunkt in der unten erweiterten Lemmazeichengestaltangabe keinen eigenen strukturellen Ort aufweisen. Vielmehr erscheint er in „**Spezialist**“ wie im konkreten Text des Wörterbuchartikels! Will man bei der Darstellung von Artikelmikrostrukturen nicht nur die artikelinterne Angabedistribution und gegebenenfalls die Distribution der Angabetexte explizit berücksichtigen, sondern (falls vorhanden) auch die der funktionalen Angabezusätze, muss man von reinen zu hybriden Mikrostrukturen übergehen. Am Beispiel von wa₅ sei dies nun zunächst stellvertretend für alle unten erweiternden funktionalen Angabezusätze erläutert, ohne dass alle formalen Details berücksichtigt werden.

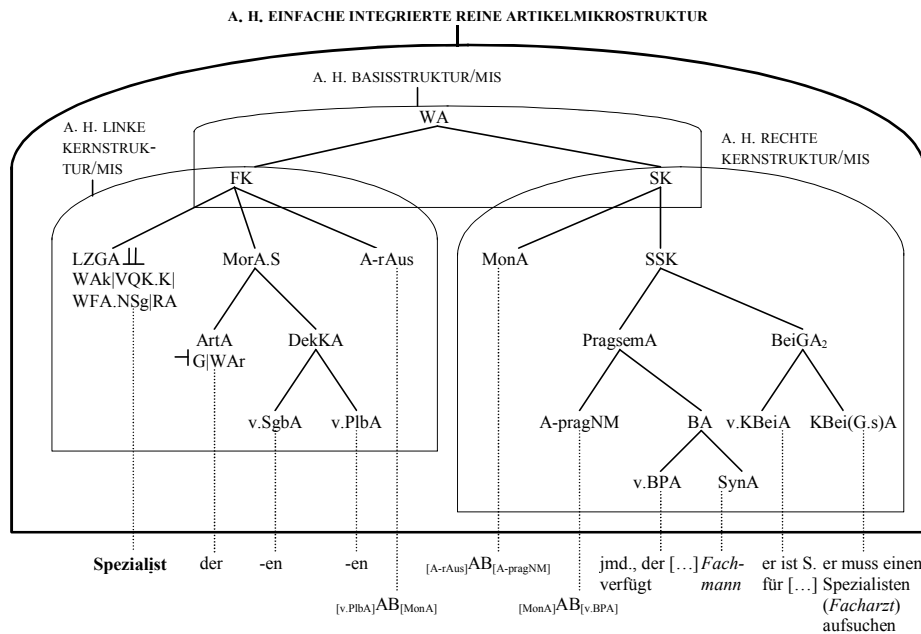


Abb. 3: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen einfachen integrierten reinen Artikelmikrostruktur, die wa₅ in Abb. 1 aufweist

Abkürzungen: WAK|VQK.K = Wortakzentkennzeichnung, zugleich Vokalquantitätskennzeichnung zur Kürze; BeiGA₂ = Beispielgruppenangabe, die aus zwei Beispielangaben besteht; KBei(G.s)A = um ein semantisches Binnenglossat (G.s) binnerweiterter Kompetenzbeispielangabe; BA = Bedeutungsangabe; SynA = Synonymangabe

Im Folgenden betrachten wir zunächst die unten erweiterte Lemmzeichengestaltangabe „**Spezialist**“ aus wa₅. Sie gehört zu den einfach erweiterten elementaren Angaben und bei diesen zu den unten erweiterten (vgl. Wiegand 2005: 289). Während nichterweiterte elementare Angaben, wie z.B. die Artikelangabe „der“ in wa₅, keine interne Struktur aufweisen, ist dies bei erweiterten elementaren Angaben anders: Ihnen kann eine interne Struktur zugewiesen werden. Deren Elemente sind bei oben oder unten erweiterten elementaren Angaben durch eine Anwendung der Methode der funktionalen segmentativen Isolierung erhältlich. Bei der Anwendung dieser Methode werden horizontale Segmentationschnitte so angelegt, dass die oben oder unten erweiternden funktionalen Angabezusätze von der elementaren Angabe abgetrennt werden. Bei der unten erweiterten Lemmzeichengestaltangabe „**Spezialist**“ bedeutet dies, dass das Segmentationsergebnis aus dem Segment „**Spezialist**“ und dem Unterpunkt besteht; es ist zu beachten, dass beide Segmente Teile von „**Spezialist**“ sind. Um die genaue Position des Unterpunktes festlegen zu können, wird

„Spezialist“ nichtfunktional-positional segmentiert, und zwar so, dass sich folgende Angabeformsegmente ergeben: **Spezial | i | st** (mit „|“ als Segmentationsfugenmarkierung) „Spezial“ ist das vordere, „i“ ist das mittlere und „st“ ist das hintere Angabeformsegment. Es gilt die Aussage: **Spezial < i < st** (mit „<“ für *geht voraus*). Mit Angabeformsegmenten werden keine bedeutungstragenden Einheiten genannt; dadurch unterscheiden sie sich von Angabesegmenten, mit denen solche Einheiten genannt werden.

Um die interne Angabestruktur von „Spezialist“ angeben zu können, gehen wir im Folgenden schrittweise vor; wir bilden zunächst folgende Trägermenge $M_{Ar}^k(\text{Spezialist}) = \{i, \bullet\}$; diese Trägermenge enthält also zwei konkrete (k) Elemente: das mittlere Angabeformsegment „i“ und den Unterpunkt. Auf $M_{Ar}^k(\text{Spezialist})$ definieren wir nun eine zweistellige irreflexive und damit asymmetrische sowie transitive Relation $R_{Ar}^k(\text{Spezialist})$ vom Typ der (textarchitektonischen) *unterhalb*-Relation, zu welcher der Relationsterm *x ist unterhalb von y* gehört, mit „x“ als Variable für funktionale Angabezusätze und „y“ als Variable für deren Bezugsadresse. Die Relation $R_{Ar}^k(\text{Spezialist})$ prägt auf der Trägermenge eine Struktur, die zu den vertikalen Angabearchitekturen zählt, da Strukturen, deren strukturprägende Relationen *oberhalb*- oder *unterhalb*-Relationen und/oder *links-von*- und *rechts-von*-Relationen sind, *Architekturen* heißen (vgl. z.B. Wiegand 2001: 191ff). Vertikale Angabearchitekturen lassen sich unter Verwendung der gleichen formalen Darstellungsmittel darstellen wie vertikale Textarchitekturen von Wörterbuchartikeln (vgl. zu diesen z.B. Bergenholtz, Tarp und Wiegand 1999: 1791ff u. Wiegand 2001: 191ff). Das allgemeine Angabearchitekturbild für die unten erweiterte Lemmazeichengestaltung aus wa₅ findet sich in Abb.4.

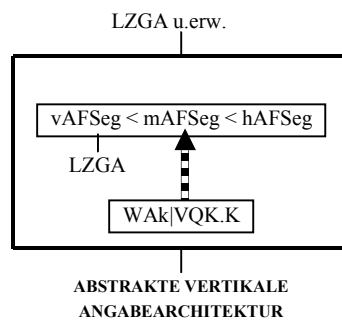


Abb. 4: Allgemeines Angabearchitekturbild zur abstrakten vertikalen Angabearchitektur der unten erweiterten Lemmazeichengestaltung aus wa₅

Abkürzungen: vAFSeg = vorderes Angabeformsegment; mAFSeg = mittleres Angabeformsegment; hAFSeg = hinteres Angabeformsegment; u. erw. = unten erweitert; „x \dashrightarrow y“ bedeutet soviel wie *x ist unterhalb von y*, wenn der Pfeil nach oben zeigt

Im Folgenden erweitern wir die Trägermenge $M_{Ar}^k(\text{Spezialist})$ um „Spezial“, das vordere Angabeformsegment, und weiterhin um „st“, das hintere Angabeformsegment, sowie um die nichterweiterte Lemmazeichengestaltangabe „Spezialist“ und schließlich noch um die unten erweiterte Lemmazeichengestaltangabe „Spezialist“, so dass wir als erweiterte elementenheterogene Trägermenge — sie heiße $M_{hyAnS}^k(\text{Spezialist})$ — eine Menge mit der Mächtigkeit $|6|$ erhalten, die wie folgt angegeben werden kann:

$$M_{hyAnS}^k(\text{Spezialist}) = \{i, \bullet, st, \text{Spezial}, \text{Spezialist}, \text{Spezialist}\}.$$

Auf dieser elementenheterogenen Trägermenge, die drei Angabeformsegmente, einen funktionalen Angabezusatz und eine nichterweiterte sowie eine unten erweiterte Angabe als Elemente enthält, definieren wir die drei strukturprägenden Relationen, deren zugehörige Relationstypen wir bereits kennen, und zwar zunächst eine Relation vom Typ der Präzedenzrelation — sie heiße $R_p^k(\text{Spezialist})$ — mit dem Relationsterm x geht voraus y , mit „ x “ und „ y “ als Variablen für Textsegmente. Es gilt $R_p^k(\text{Spezialist}) \subseteq M_{hyAnS}^k(\text{Spezialist}) \times M_{hyAnS}^k(\text{Spezialist})$. — Weiterhin wird auf $M_{hyAnS}^k(\text{Spezialist})$ eine Relation vom Typ der partitiven Relation — sie heiße $R_{part}^k(\text{Spezialist})$ — definiert, mit dem Relationsterm u ist ein Teil von v mit „ u “ und „ v “ als Variable für Textsegmente. Es gelten:

$$R_{part}^k(\text{Spezialist}) \subseteq M_{hyAnS}^k(\text{Spezialist}) \times M_{hyAnS}^k(\text{Spezialist}) \text{ sowie} \\ R_{part}^k(\text{Spezialist}) \cup R_p^k(\text{Spezialist}) = \emptyset.$$

— Schließlich definieren wir auf $M_{hyAnS}^k(\text{Spezialist})$ eine Relation vom Typ der textarchitektonischen *unterhalb*-Relation — sie heiße $R_{unt}^k(\text{Spezialist})$ — mit dem Relationsterm r ist unterhalb von s , mit „ r “ als Variable für funktionale Angabezusätze und „ s “ als Variable für deren Bezugsadressen. Es gelten:

$$R_{unt}^k(\text{Spezialist}) = M_{hyAnS}^k(\text{Spezialist}) \times M_{hyAnS}^k(\text{Spezialist}); R_{unt}^k(\text{Spezialist}) \cap R_p^k(\text{Spezialist}) = \emptyset; R_{unt}^k(\text{Spezialist}) \cap R_{part}^k(\text{Spezialist}) = \emptyset.$$

Als Ergebnis aller genannten Operationen erhalten wir die konkrete hierarchische architektonisch angereicherte Angabestruktur der unten erweiterten Lemmazeichengestaltangabe aus wa_5 ; diese ist in Abb. 5 (1) dargestellt; in Abb. 5 (2) findet sich die zugehörige isomorphe abstrakte hierarchische architektonisch angereicherte Angabestruktur, die dadurch erhältlich ist, dass man auf der elementenheterogenen Trägermenge

$$M_{hyAnS}^a(\text{Spezialist}) = \{mAFSeg, WAK | VQK.K, vAFSeg, hAFSeg, LZGA \sqcup WAK | VQK.K\}.$$

drei Relationen definiert, die zu demselben Relationstyp gehören wie die drei auf $M_{hyAnS}^k(\text{Spezialist})$ definierten Relationen.

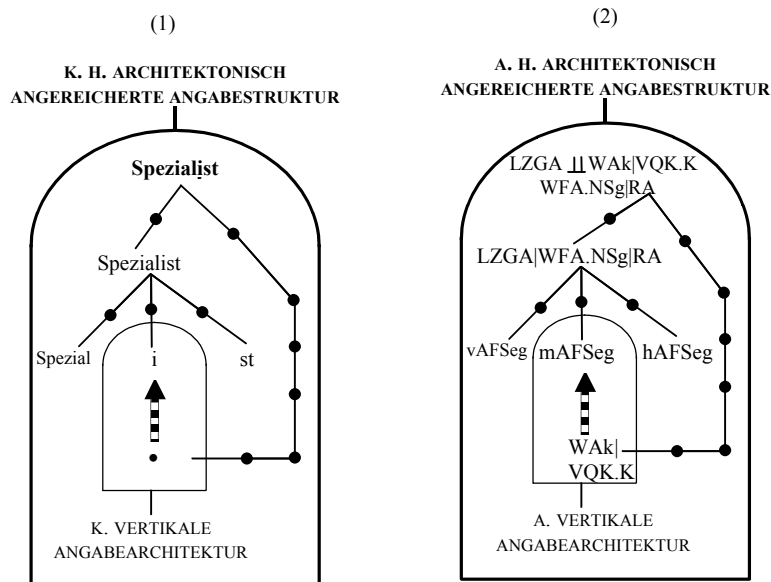


Abb. 5: Einfach kommentierte und mit architektonischen Komponenten angereicherte Baumgraphen zur konkreten (1) und zur isomorphen abstrakten hierarchischen Angabestruktur (2) der unten erweiterten Lemmazeichengestaltung aus was
Abkürzungen: K = KONKRETE; *Darstellungskonventionen:* „x \dashrightarrow y“ bedeutet soviel wie x ist unterhalb von y; „u \dashv v“ bedeutet soviel wie u ist ein Teil von v, mit „u“ als Variable für Textsegmente, die keine Textkonstituenten sind

Die beiden in Abb. 5 präsentierten Baumgraphen lassen sich zu einem Strukturgraphen zusammenfassen; dieser findet sich in Abb. 6.

Weiter oben wurde festgelegt, dass zu jedem Wörterbuchartikel, in dem mindestens ein funktionaler Angabezusatz gegeben ist, sowohl eine reine als auch mindestens eine hybride Artikelmikrostruktur angegeben werden kann. Der Übergang von einer reinen zu einer hybriden Struktur kann auch ausschließlich auf der Ebene der Strukturdarstellung erfolgen, da es sich um strikt formale Strukturdarstellungen handelt. Demgemäß wird nun der Strukturgraph in Abb. 6 in den Strukturgraph in Abb. 3 integriert, so dass sich der Strukturgraph von wa₅ ergibt; dieser findet sich in Abb. 7, wobei auf die vollständige Wiedergabe der Kommentarstruktur des semantischen Kommentars verzichtet wurde, weil der funktionale Angabezusatz im semantischen Kommentar von wa₅ noch nicht analysiert ist.

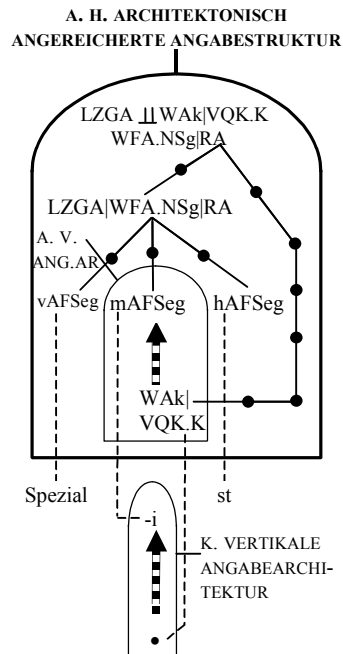


Abb. 6: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen Angabestruktur der unten erweiterten Lemmazichengestaltangabe aus wa_3

Abkürzungen: A. V. ANG.AR = ABSTRAKTE VERTIKALE ANGABEARCHITEKTUR; *Darstellungskonventionen* wie in Abb. 5; *Darstellungskonventionen:* „x---Y“ bedeutet (von unten nach oben gelesen) soviel wie x ist ein Element von Y ($= x \in Y$)

Bisher wurden hybride Angabestrukturen behandelt, die zu dem Typ der hierarchischen architektonisch angereicherten Angabestrukturen gehören, der ein Untertyp des Typs der hierarchischen hybriden konstituentenlosen Angabestruktur ist; der zuletzt genannte Strukturtyp weist einen weiteren Untertyp auf, nämlich den Typ der hierarchischen binnenerweiterungsbedingten Angabestruktur (vgl. Wiegand 2007: 201ff u. Abb. 20). Handelt es sich, wie z.B. in wa_5 , bei den binnenerweiternden funktionalen Angabezusätzen um Binnenglossate, wird von *hybriden binnenglossatbedingten Angabestrukturen* gesprochen und verkürzt von *hybriden glossatbedingten Angabestrukturen*. Der verkürzte Terminus ist ausreichend deutlich, da postglossierte Angaben stets zu den nicht-elementaren Angaben gehören, so dass „postglossatbedingte Angabestrukturen“ nicht auftreten können. Im Folgenden betrachten wir glossatbedingte Angabestrukturen.

A. H. INTERNLEMMATISCHE ARCHITEKTONISCH ANGEREICHERTE ARTIKELMIKROSTRUKTUR

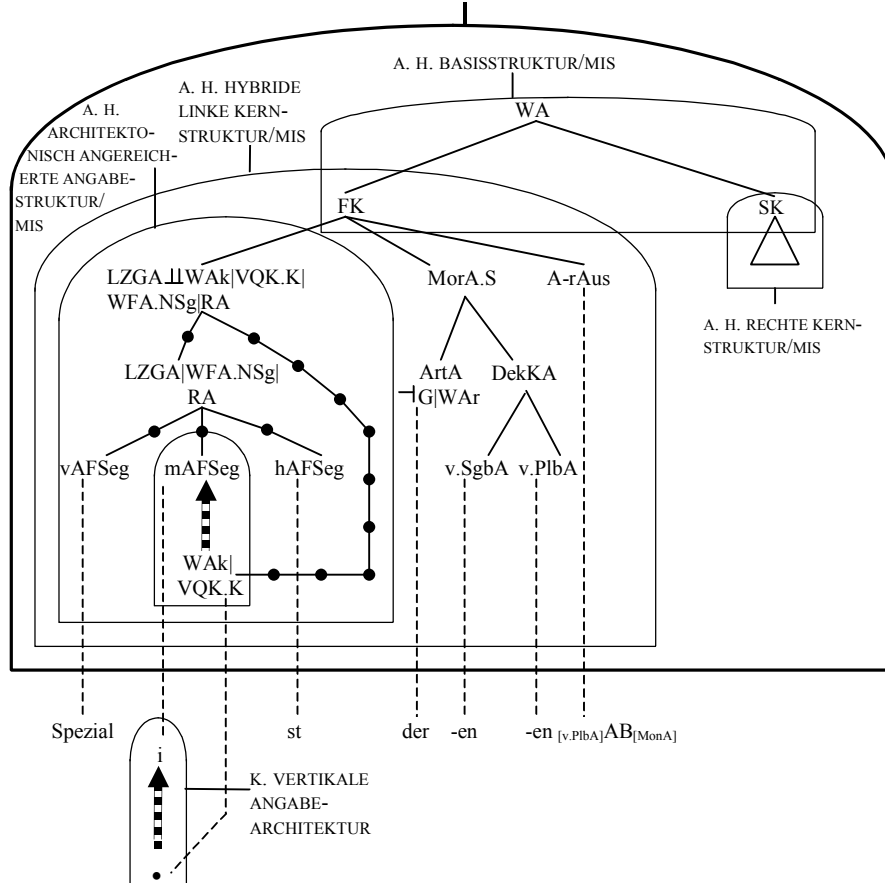


Abb. 7: Nichtvollständig ausgeführter, einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen internlemmatisch architektonisch angereicherten Artikelmikrostruktur, die wa₃ in Abb. 1 aufweist

Die Angabe „er muß einen Spezialisten (*Facharzt*) aufsuchen“ (= a₁) aus wa₅ in Abb. 1 ist eine um ein semantisches Binnenglossat (G.s) erweiterte Kompetenzbeispielangabe; sie erhält den Angabennamen a₁; so dass gilt a₁ ∈ KBei(G.s)A. Die Angabe a₁ gehört zu den elementaren Angaben, da a₁ funktional-positional nicht segmentierbar ist; a₁ weist mithin keine (konkrete und abstrakte) Angabekonstituenten- oder (konkrete und abstrakte) Angabemikrostrukturen als deren Teilstrukturen auf. Dennoch weist die Angabe a₁ eine Struktur auf, die ein Benutzer, der anhand dieser Angabe eine Information gewinnen möchte, auch intuitiv erkennen muss. Diese konkrete (k) hierarchische glossatbedingte

(g) Angabestruktur (AnS) ist wie folgt erhältlich: Man bildet zunächst eine elementenheterogene Trägermenge — sie heiße $M_{gAnS}^k(a_1)$ — die als Elemente diejenigen Segmente enthält, die durch eine Anwendung der Methode der nichtfunktionalen-positionalen Segmentationen auf a_1 erhältlich sind. $M_{gAnS}^k(a_1)$ kann wie folgt angegeben werden:

$$M_{gAnS}^k(a_1) = \{b_1, c_1, d_1, e_1, f_1, a_1\}.$$

Die Elemente der Trägermenge sind wie folgt erklärt:

- b_1 = er muß einen Spezialisten; $b_1 \in vASeg$ (= vorderes Angabesegment).
- c_1 = (; $c_1 \in vZOZ$ (= vorderes Zusammenordnungszeichen).
- d_1 = *Facharzt*; $d_1 \in G.s$ (= semantisches Binnenglossat).
- e_1 =); $e_1 \in hZOZ$; (= hinteres Zusammenordnungszeichen).
- f_1 = aufsuchen; $f_1 \in hASeg$ (= hinteres Angabesegment).

Nach der Bildung der Trägermenge $M_{gAnS}^k(a_1)$ definiert man auf dieser zwei strukturprägende Relationen, und zwar eine vom Typ der Präzedenzrelation — sie heiße $R_p^k(a_1)$ — mit dem Relationsterm *x geht voraus y* (mit „x“ und „y“ als Variablen für Textsegmente), sowie eine vom Typ der partitiven Relation — sie heiße $R_{part}^k(a_1)$ — mit dem Relationsterm *x ist ein Teil von y* (mit „x“ als Variable für Textsegmente und „y“ als Variable für binnenerweiterte Angaben). Es gelten: $R_p^k(a_1) \subseteq M_{gAnS}^k(a_1) \times M_{gAnS}^k(a_1)$; $R_{part}^k(a_1) \subseteq M_{gAnS}^k(a_1) \times M_{gAnS}^k(a_1)$; $R_p^k(a_1) \cap R_{part}^k(a_1) = \emptyset$. Die konkrete hierarchische glossatbedingte Angabestruktur von a_1 , die sich nach der Ausführung der genannten Prozeduren ergibt, ist in Abb. 8 (1) dargestellt. Definiert man auf der Trägermenge für die zugehörige abstrakte (a) hierarchische glossatbedingte (g) Angabestruktur (AnS)

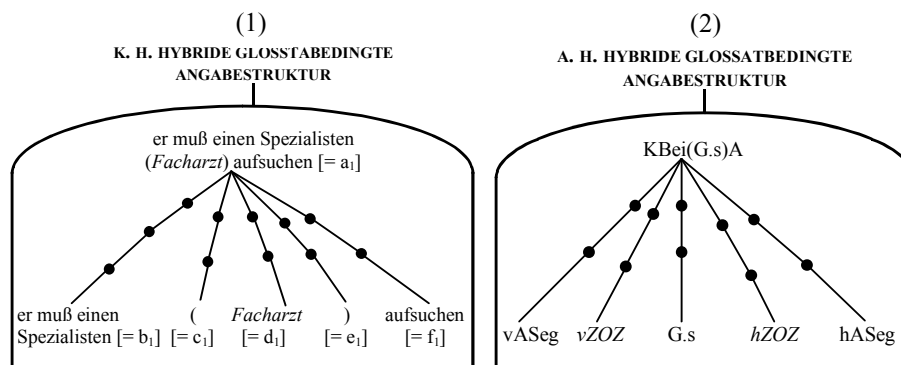


Abb. 8: Erweitert kommentierter Baumgraph zur konkreten (1) und einfach kommentierter Baumgraph zur isomorphen abstrakten (2) hierarchischen hybriden glossatbedingten Angabestruktur, die die binnenerweiterte Kompetenzbeispielangabe aus wa_5 in Abb. 1 aufweist

— sie heie $M_{gAnS}^a(a_1)$ — ($M_{gAnS}^a(a_1) = \{M_{gAnS}^a(a_1) = \{vASeg, vZOZ, G.s, hZOZ, hASeg, KBei(G.s)A\}$) zwei Relationen $R_p^a(a_1)$ und $R_{part}^a(a_1)$, die den auf $M_{gAnS}^k(a_1)$ definierten entsprechen, dann erhlt man die abstrakte hierarchische glossatbedingte Angabestruktur von a_1 , die in Abb. 8 dargestellt ist.

Wie die beiden Baumgraphen in Abb. 5, so lassen sich auch die beiden Baumgraphen in Abb. 8 zu einem Strukturgraphen zusammenfassen; dieser findet sich in Abb. 9.

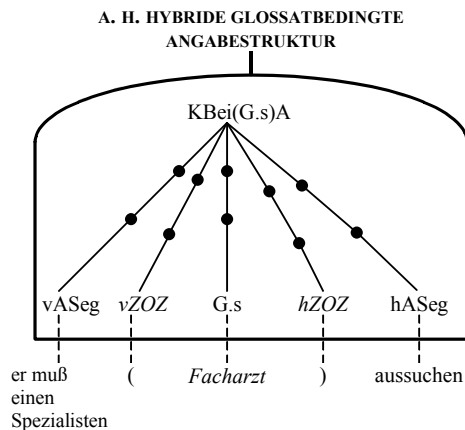


Abb. 9: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen hybriden glossatbedingten Angabestruktur, die die binnenerweiterte Kompetenzbeispielangabe in was in Abb. 1 aufweist

Wie aus den Abb. 8 und 9 und aus den Textzeilen zur Bildung von glossatbedingten Angabestrukturen hervorgeht, werden bei diesen Strukturen die nichttypographischen Angabestrukturanzeiger, die von nichttypographischen Angabemikrostrukturanzeigern unterschieden werden mssen (vgl. WLWF-1 2010) bercksichtigt. Dies bedeutet, dass glossatbedingte Angabestrukturen nicht als Teilstrukturen von hierarchischen Artikelmikrostrukturen auftreten, da in letzteren die nichttypographischen Strukturanzeiger keine Bercksichtigung finden. Vielmehr treten hierarchische glossatbedingte Angabestrukturen als Teilstrukturen von hierarchischen Artikelkonstituentenstrukturen auf, bei denen die nichttypographischen Strukturanzeiger als Textkonstituenten bercksichtigt werden (vgl. Abb. 14). Im Folgenden wird die in Abb. 9 in konkreter und abstrakter Ausprgung dargestellte hierarchische glossatbedingte Angabestruktur in den semantischen Kommentar der Artikelkonstituentenstruktur von wa_5 in Abb. 1 strukturell integriert. Das Ergebnis dieser Operation findet sich in Abb. 10.

Will man die hierarchische Struktur von binnenerweiterten Angaben bei der Darstellung von hybriden Artikelmikrostrukturen explizit bercksichtigen, bentigt man hierarchische Angabestrukturen, in denen die nichttypographischen Angabestrukturanzeiger nicht bercksichtigt werden. Solche Angabe-

strukturen heißen (nach Wiegand 2007: 201f u. 2009a: 80f) *hierarchische hybride minimierte glossatbedingte Angabestrukturen*. Sie sind auf zweierlei Weise erhältlich. Verfügt man bereits — wie hier — über glossatbedingte Angabestrukturen, dann kann man die Trägermengen um die entsprechenden Elemente verringern. Demgemäß müssen *erstens* die Trägermengen für die konkreten hierarchischen glossatbedingten Angabestrukturen um die konkreten Elemente verringert werden; die Trägermengen für die abstrakten hierarchischen glossatbedingten Angabestrukturen müssen entsprechend um die Elemente verringert werden, die Klassen von nichttypographischen Strukturanzeigern sind.

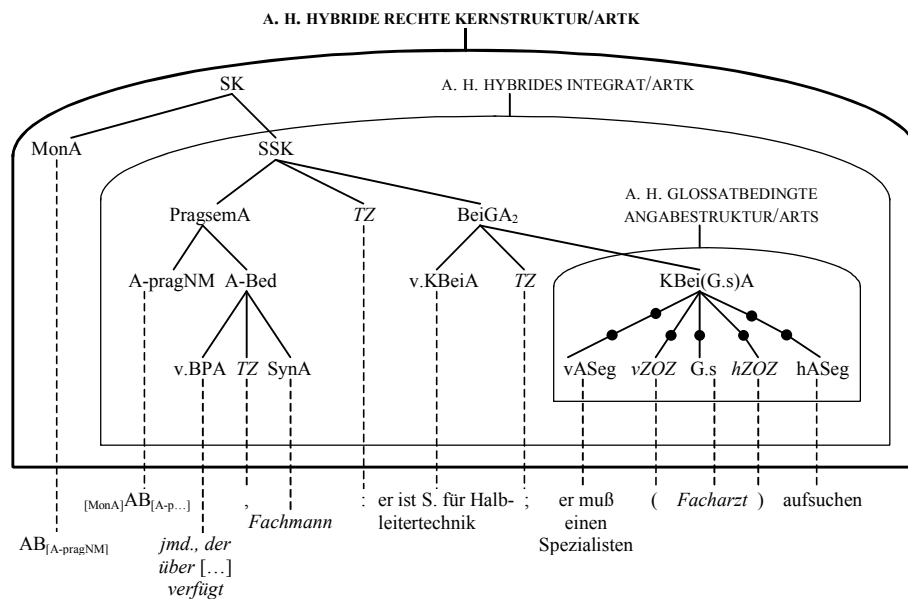


Abb. 10: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen hybriden Kernstruktur der hybriden Artikelkonstituentenstruktur, die wa₅ in Abb. 1 aufweist

Abkürzungen: BeiGA₂ = Beispielgruppenangabe, die aus zwei Beispielangaben besteht; TZ = Trennzeichen; vZOZ = vorderes Angabesegment; G.s = semantisches Binnenglossat; hZOZ = hinteres Zusammenordnungszeichen; hASeg = hinteres Angabesegment; ARTK = ARTIKELKONSTITUENTENSTRUKTUR; KERNSTRUKTUR/ARTK ist zu lesen wie *Kernstruktur als Substruktur der Artikelkonstituentenstruktur*

Zweitens müssen die auf den beiden Trägermengen definierten Relationen auf die verringerte Trägermenge eingeschränkt werden, so dass alle 2-Tupel getilgt werden, in denen als Koordinate entweder (bei den konkreten Strukturen) ein nichttypographischer Strukturanzeiger auftritt, oder (bei den abstrakten Strukturen) eine Klasse von nichttypographischen Strukturanzeigern. Bei Angabe a₁ = „er muß einen Spezialisten (*Facharzt*) aufsuchen“ aus wa₅ in Abb. 1 ergibt sich dann das Folgende:

- (1) Die Trägermenge für die konkrete Struktur, nämlich $M_{gAnS}^k(a_1)$, wird um die Elemente c_1 und e_1 (vgl. Abb. 8 (1)) vermindert, so dass sich die verminderte (v) Trägermenge $vM_{gAnS}^k(a_1) = \{b_1, d_1, f_1, a_1\}$ ergibt. Die Relation $R_p^k(a_1)$ wird so eingeschränkt, dass folgende ihrer 2-Tupel, die c_1 oder e_1 als Koordinaten aufweisen, eliminiert werden: $\langle b_1, c_1 \rangle$, $\langle b_1, e_1 \rangle$, $\langle c_1, d_1 \rangle$, $\langle c_1, e_1 \rangle$, $\langle c_1, f_1 \rangle$, $\langle d_1, e_1 \rangle$, und $\langle e_1, f_1 \rangle$. Die eingeschränkte (e) Relation $eR_p^k(a_1)$ kann dann wie folgt angegeben werden: $eR_p^k(wa_1) = \{\langle b_1, d_1 \rangle, \langle b_1, f_1 \rangle, \langle d_1, f_1 \rangle\}$ (vgl. Abb. 11). Die Relation $R_{part}^k(a_1)$ wird ebenfalls eingeschränkt, und zwar so, dass folgende ihrer 2-Tupel, die c_1 oder e_1 als Koordinaten aufweisen, eliminiert werden: $\langle c_1, a_1 \rangle$, $\langle e_1, a_1 \rangle$. Die eingeschränkte (e) Relation kann dann wie folgt angegeben werden: $eR_{part}^k(a_1) = \{\langle b_1, a_1 \rangle, \langle d_1, a_1 \rangle, \langle f_1, a_1 \rangle\}$.
- (2) Die Trägermenge für die abstrakte (a) Struktur, nämlich $M_{gAnS}^a(a_1)$, wird um die Elemente $vZOZ$ und $hZOZ$ vermindert, so dass sich die verminderte (v) Trägermenge $vM_{gAnS}^a(a_1) = \{vASeg, G.s, hASeg\}$ ergibt. Die Relation $R_p^a(a_1)$ wird analog zu $R_p^k(a_1)$ eingeschränkt und kann demgemäß wie folgt angegeben werden: $eR_p^a(a_1) = \{\langle vASeg, G.s \rangle, \langle vASeg, hASeg \rangle, \langle G.s, hASeg \rangle\}$. Schließlich wird noch $R_{part}^a(a_1)$ analog zu $R_{part}^k(a_1)$ eingeschränkt und kann demgemäß wie folgt angegeben werden: $eR_{part}^a(a_1) = \{\langle vASeg, KBei(G.s)A \rangle, \langle G.s, KBei(G.s)A \rangle, \langle hASeg, KBei(G.s)A \rangle\}$.

Nach Ausführung der skizzierten Operationen ergeben sich die konkrete hierarchische minimierte glossatbedingte Angabestruktur sowie die zugehörige isomorphe abstrakte hierarchische minimierte glossatbedingte Angabestruktur, die a_1 , die um eine semantisches Binnenglossat erweiterte Kompetenzbeispielangabe „er muß einen Spezialisten (*Facharzt*) aufsuchen“ aus wa_5 in Abb. 1 aufweist; die beiden minimierten Strukturen sind mit dem Strukturgraphen in Abb. 11 dargestellt.

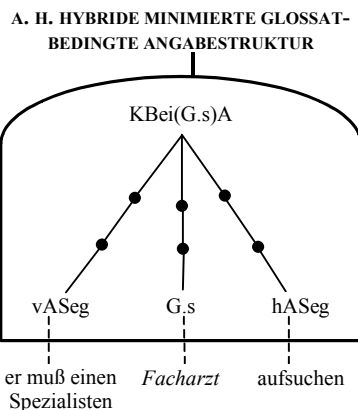


Abb. 11: Einfach kommentierter Strukturgraph zur konkreten (und isomorphen abstrakten) minimierten glossatbedingten Angabestruktur, die wa_5 in Abb. 1 aufweist

Der skizzierte Übergang von hierarchischen glossatbedingten zu hierarchischen minimierten glossatbedingten Angabestrukturen lässt sich auch auf der Ebene der formalen Strukturdarstellung durchführen, indem man die Kanten, die zu den Knoten führen, die nichttypographische Angabestrukturanzeiger repräsentieren, sowie die zugehörigen Knotenetikette tilgt und die einfache Kommentierung sinngemäß ändert.

Wir verfügen nun über alle Substrukturen, um die konkrete (und isomorphe abstrakte) hierarchische internlemmatisch architektonisch angereicherte Artikelmikrostruktur mit minimierter glossatbedingter Teilstruktur, die wa₅ in Abb. 1 aufweist, in Abb. 12 darstellen zu können.

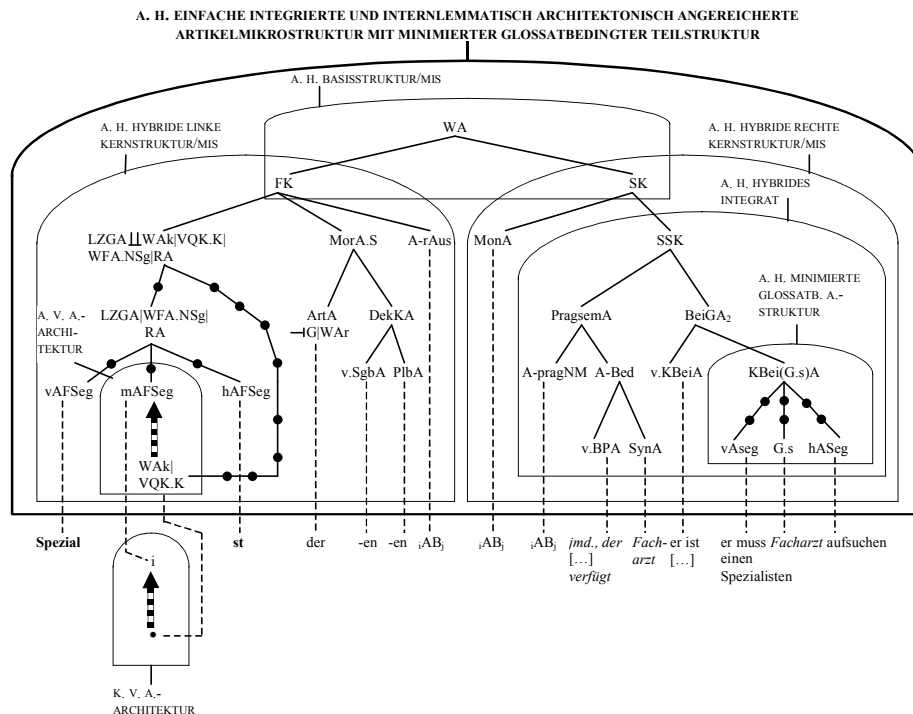


Abb. 12: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen einfachen integrierten und internlemmatisch architektonisch angereicherten Artikelmikrostruktur mit minimierter glossatbedingter Teilstruktur

Abkürzungen: vAFSeg = vorderes Angabeformsegment; mAFSeg = mittleres Angabeformsegment; hAFSeg = hinteres Angabeformsegment; A-rAus = Angabe zur regelmäßigen Aussprache; MonA = Monosemieangabe; A-pragNM = Angabe zur pragmatischen Nullmarkierung; A-Bed = Angabe zur Bedeutung; v.BPA = verdichtete Bedeutungsparaphrasenangabe; BeiGA₂ = Beispielgruppenangabe, die aus zwei Beispielangaben (unterschiedlichen Typs) besteht; v.KBeiA = verdichtete Kompetenzbeispielangabe; KBei(G.s)A = um ein

semantisches Binnenglossat erweitere Kompetenzbeispielangabe; vASeg = vorderes Angabesegment; G.s = semantisches Binnenglossat; hASeg = hinteres Angabesegment; ;AB_j = Angabeblank mit Nachbarschaftsvariablen ohne Belegung. A.V.A.-ARCHITEKTUR = ABSTRAKT VERTIKALE ANGABEARCHITEKTUR; K.V.A.-ARCHITEKTUR = KONKRETE VERTIKALE ANGABEARCHITEKTUR

Im Folgenden betrachten wir wa₆ in Abb. 1. Der Artikel weist eine konkrete (und isomorphe abstrakte) hierarchische rechtserweiterte integrierte hybride Artikelmikrostruktur mit minimierter doppelglossatbedingter Teilstruktur auf. Das Doppelglossat findet sich in der verdichteten Kompetenzbeispielangabe „das hat viel S. (*Mühe, Anstrengung*) gekostet“. Ein Doppelglossat ist ein Glossat, das aus zwei aufeinanderfolgenden gleichadressierten Binnenglossaten besteht, zwischen denen ein Trennzeichen steht. Die Glossate sind hier nicht nur gleichadressiert, sondern auch klassengleich: Es handelt sich hier um zwei semantische Binnenglossate. Für Doppelglossate mit zwei klassengleichen Glossaten gibt es zwei Analysemöglichkeiten, zu denen zwei unterschiedliche Strukturkonzeptionen gehören.

- (1) Die erste Analysemöglichkeit wurde in Wiegand (2009a: 79f) vorgeführt; sie basiert darauf, dass das Konzept der homosegmentären Textsegmente nur für klassengleiche Angaben gelten soll. Nach diesem Strukturkonzept ist z.B. die Angabe „*Knabe, Junge*“ in wa₇ in Abb. 14 eine (nichtelementare) zweifach homosegmentäre Synonymenangabe (SynA²), die in zwei (elementare) Synonymangaben (SynA) funktional-positional segmentierbar ist (vgl. Wiegand 2005: 229f). Das semantische Doppelglossat „*Mühe, Anstrengung*“ in wa₆ wird dann nicht als homosegmentäres Textsegment analysiert. Als Grund dafür kann angegeben werden, dass alle binnenerweiterungsbedingten hierarchischen Angabestrukturen als *flache hierarchische Strukturen* (i.S.v. Wiegand 2009: 254f) zu konzipieren sind. Das sind — um es einfach auszudrücken — solche textuellen Strukturen, bei deren Darstellung mittels eines geordneten Baumgraphen außer dem Wurzelknoten nur terminale Knoten auftreten, wie z.B. in Abb. 8, 9 und 11. Diese Strukturkonzeption ist vertretbar, wenn in einem Wörterbuch nur elementare Binnenglossate und nichtelementare Doppelglossate auftreten, die aus zwei klassengleichen Binnenglossaten bestehen. Treten nichtelementare Binnenglossate auf, die zu anderen Typen gehören, wie in wa₈ in Abb. 15, ist die zweite Strukturkonzeption vorzuziehen.
- (2) Die zweite Analysemöglichkeit ergibt sich, wenn man das Konzept der homosegmentären Textsegmente auch für Doppelglossate gelten lässt und damit *tiefe hierarchische Strukturen* bei den binnenerweiterungsbedingten Angabestrukturen zulässt, also solche Strukturen, bei deren Darstellung mittels eines geordneten Baumgraphen mindestens ein Kantenzug auftritt, so dass die Höhe des Baumgraphen mindestens den

Wert 2 hat. In Abb. 13 ist die konkrete (und isomorphe abstrakte) hierarchische minimierte doppelglossatbedingte Angabestruktur der um ein semantisches Doppelglossat binnerweiterter verdichteten Kompetenzbeispielangabe des metaphorischen Wortgebrauchs „das hat viel S. (*Mühe, Anstrengung*) gekostet“ aus wa₆ sowohl als flache (1) als auch als tiefe (2) Struktur dargestellt.

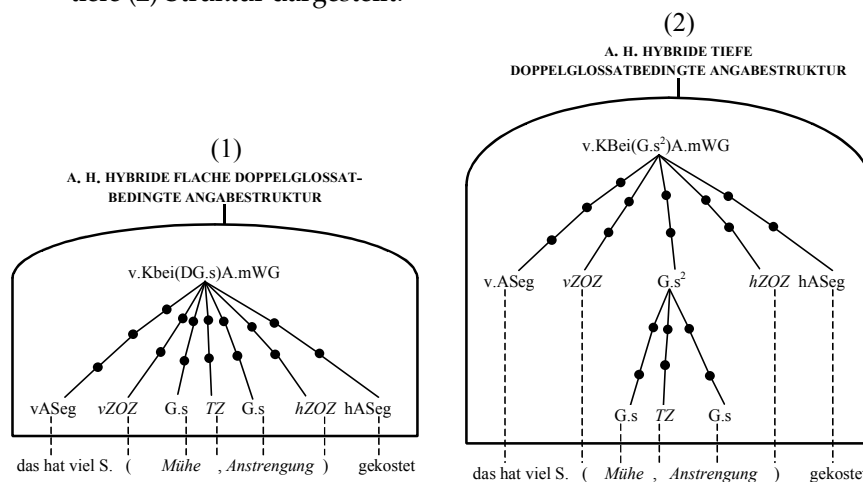


Abb. 13: Einfach kommentierte Strukturgraphen zur abstrakten (und isomorphen konkreten) hierarchischen doppelglossatbedingten Angabestruktur, konzipiert als flache Struktur (1) und als tiefe Struktur (2)

Abkürzungen: DG.s = semantisches Doppelglossat; G.s² = zweifach homosegmentäres semantisches Binnenglossat (= semantisches Doppelglossat, verstanden als homosegmentäres Textsegment); v.KBei(DG.s)A.mWG = um ein semantisches Doppelglossat binnerweiterter verdichtete Kompetenzbeispielangabe des metaphorischen Wortgebrauchs

Zu dem Segmentationsprozess als ersten Prozessabschnitt des Strukturbildungsprozesses, der zu der in Abb. 13 (2) dargestellten tiefen Struktur geführt hat, sei das Folgende nachgetragen: Es gehört zu den Korrektheitsbedingungen bei der Anwendung der Methode der nichtfunktionalen-positionalen Segmentierung, dass ihre Anwendung zu mindestens einem nichtfunktionalen Textsegment führt; es gehört aber nicht zu den Korrektheitsbedingungen, dass alle erhältlichen Textsegmente nichtfunktionale Textsegmente sind. Demgemäß läuft die Segmentierung der um ein semantisches Doppelglossat binnerweiterter verdichteten Kompetenzbeispielangabe des metaphorischen Wortgebrauchs (v.KBei(DG.s)A.mWG) wie folgt ab: Im ersten Schritt wird „das hat viel S. (*Mühe, Anstrengung*) gekostet“ nichtfunktional-positional segmentiert, so dass fünf Textsegmente gegeben sind (das hat viel S. | (| *Mühe, Anstrengung* |) | gekostet, mit „|“ als Segmentationsfugenmarkierung). Da diese fünf Textsegmente nicht das Ergebnis einer Anwendung der funktional-positionalen

Segmentation sind, sind sie keine Textkonstituenten. Zwei Textsegmente sind nichtfunktional, nämlich „das hat viel S.“ \in vASeg“ und „gekostet“ \in hASeg. Der diskontinuierliche angabeinterne Strukturanzeiger, das Zusammenordnungszeichen „()“, das durch ein Zeichen-Tupel realisiert ist, ist dagegen ein funktionales Textsegment. Das Doppelglossat ist ein nichtelementares funktionales Textsegment ohne Textkonstituentenstatus, da es — wie bereits gesagt — nicht aus einer Anwendung einer der Methodenvarianten der Methode der funktional-positionalen Segmentation hervorgegangen ist. Auf das isolierte Doppelglossat „Mühe, Anstrengung“ lässt sich nun allerdings die Methode der exhaustiven funktional-positionalen Segmentation anwenden, da das Doppelglossat restfrei in drei funktionale Textsegmente segmentiert werden kann. Daher ergibt sich nun die Frage, ob diese drei Textsegmente Textkonstituentenstatus aufweisen oder nicht. Diese Frage lässt sich durch Rückgriff auf die zentrale Korrektheitsbedingung für die Anwendung der Methode der funktional-positionalen Segmentation beantworten, die lautet: „Durch die Anwendung der Methode der funktional-positionalen Methode auf ein Textsegment TS entstehen ausschließlich und mindestens zwei funktionale Textsegmente, die genau dann Textkonstituentenstatus haben, wenn TS eine Textkonstituente ist. Ist TS ein Textsegment ohne Textkonstituentenstatus, dann sind die durch eine Anwendung der funktional-positionalen Methode erhältlichen funktionalen Textsegmente hybride Textkonstituenten“ (vgl. Wiegand 2010). Da das Doppelglossat keine Textkonstituente ist, aber funktional positional segmentierbar, sind „Mühe“, das Komma und „Anstrengung“ hybride Textkonstituenten. Damit ist der erste Typ der hybriden textuellen Einheiten identifiziert (vgl. 7).

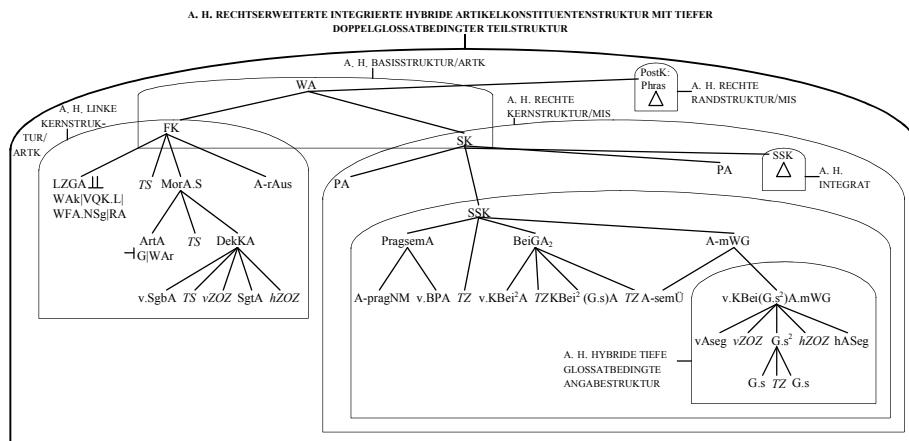


Abb. 14: Nichtvollständig ausgeführter einfach kommentierter Strukturgraph zur abstrakten hierarchischen rechtserweiterten integrierten hybriden Artikelkonstituentenstruktur mit tiefer doppelglossatbedingter Teilstruktur, die wa₆ in Abb. 1 aufweist

Abkürzungen: SgtA = Singularetantumangabe; v.KBei²A = verdichtete Kompetenzbeispielangabe, anhand derer zwei Beispiele erschlossen werden können; A.mWG = Angabe des metaphorischen Wortgebrauchs; A-semÜ = Angabe zum semantischen Übergang; v.KBeiA(G.s²)A.mWG = verdichtete Kompetenzbeispielangabe des metaphorischen Wortgebrauchs, binnenerweitert um ein zweifach homosegmentäres Glossat (= Doppelglossat)

Im Folgenden wird die in Abb. 13 (2) dargestellte abstrakte hierarchische tiefe doppelglossatbedingte Angabestruktur in die Artikelkonstituentenstruktur von wa₆ in Abb. 1 integriert. Der Strukturbildungsprozess wird nicht beschrieben, sondern es wird nur sein Ergebnis in Abb. 14 partiell dargestellt.

Zum Abschluss dieses Abschnittes findet sich in Abb. 15, um einen partiellen Überblick zu ermöglichen, ein Ausschnitt aus einer Typologie hierarchischer hybrider textueller Wörterbuchstrukturen, der auch erkennen lässt, welche Typen in diesem Beitrag nicht behandelt werden konnten und der in Abschnitt 5 ergänzt wird (vgl. Abb. 20).

Zu den Termini für hierarchische hybride Artikelkonstituenten- und für hierarchische hybride Artikelmikrostrukturen ist zu bemerken: Alle architektonisch angereicherten Textkonstituentenstrukturen sind hybrid, so dass in einem Terminus, in dem das komplexe Prädikat *textarchitektonisch angereicherte* auftritt, das Prädikat *hybrid* überflüssig ist.

5. Hybride artikelinterne Textkonstituentenstrukturen II: Angabekonstituenten- und Angabemikrostrukturen

Hybride konstituentenlose Angabestrukturen sind bereits bekannt. Die in Abb. 5 (1) u. (2), Abb. 6, Abb. 8 (1) u. (2), Abb. 9 und Abb. 11 dargestellten Strukturen von elementaren durch funktionale Angabezusätze erweiterten Angaben gehören jeweils zu Untertypen des Typs der hybriden konstituentenlosen Angabestruktur (vgl. Abb. 20). Während hybride konstituentenlose Angabestrukturen stets zu elementaren oben oder unten oder binnenerweiterten Angaben gehören, sind hybride Angabekonstituentenstrukturen und deren Teilstrukturen, die hybriden Angabenmikrostrukturen, stets Strukturen von nicht-elementaren Angaben. Auch Beispiele für diese Strukturen, nämlich bestimmte Kommentarstrukturen, sind bereits bekannt. Der semantisch Kommentar von wa₅ in Abb. 1 ist (als unmittelbare Textkonstituente des Wörterbuchartikels wa₅) eine nichtelementare Angabe mit Kommentarstatus. Die in Abb. 10 dargestellte konkrete hierarchische hybride rechte Kernstruktur ist als Teilstruktur der hierarchischen hybriden Artikelkonstituentenstruktur eine konkrete hybride Angabekonstituentenstruktur und gehört daher auch zum Typ der konkreten hierarchischen hybriden Angabekonstituentenstruktur mit flacher doppelglossatbedingter Teilstruktur (vgl. Abb. 20). Entsprechendes gilt für die zugehörigen abstrakten Strukturen. — Die in Abb. 12 dargestellte abstrakte hierarchische hybride linke Kernstruktur gehört als Teilstruktur der abstrakten hierarchischen hybriden Artikelmikrostruktur zu den abstrakten hierarchischen

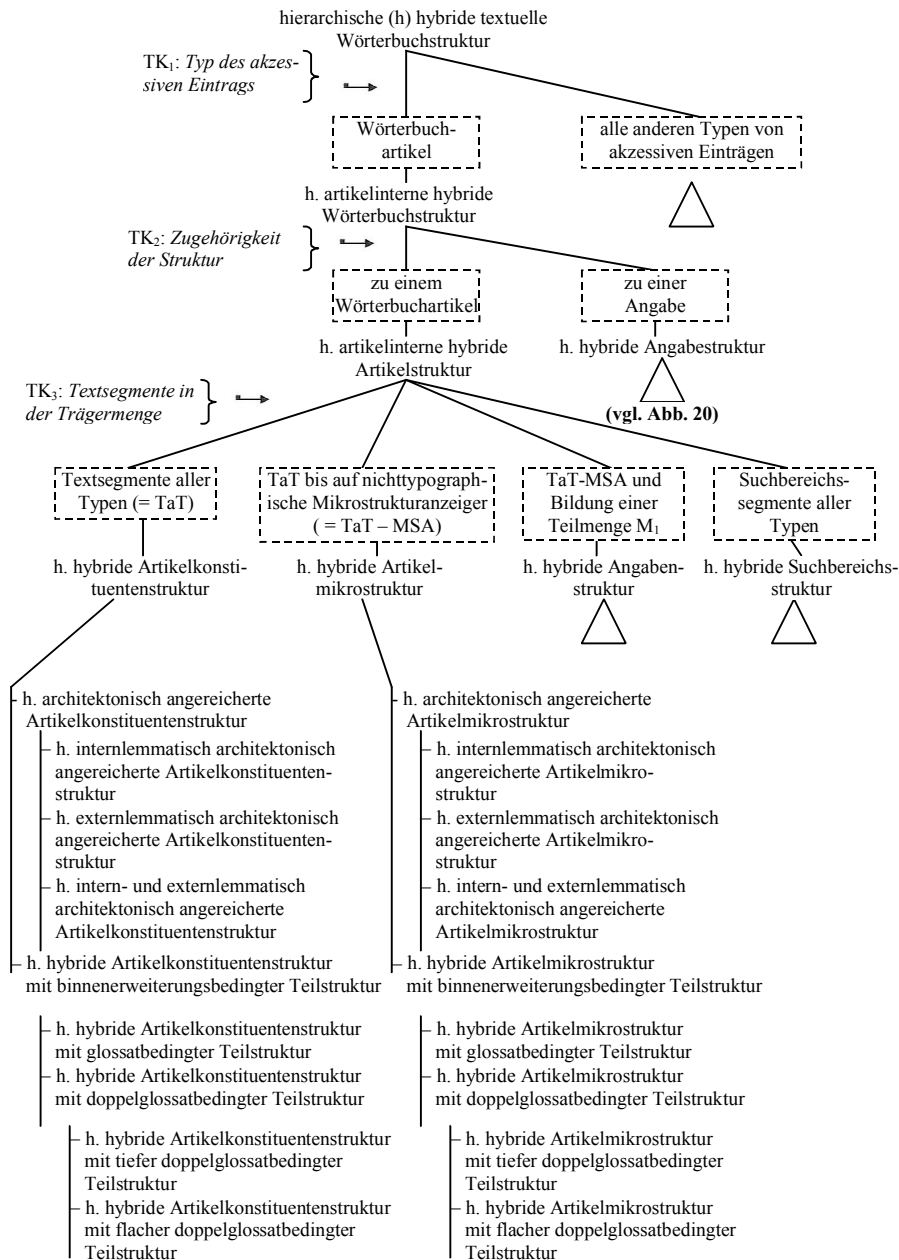


Abb. 15: Partiiell erweitert kommentierter Typologieglyph zu einem Ausschnitt aus einer Typologie hierarchischer hybrider textueller Wörterbuchstrukturen

Abkürzungen: TK = Typologiekriterium; h. = hierarchisch; TaT = Textsegmente aller Typen; TaT-MSA = Textsegmente aller Typen minus nichttypogra-

phische Mikrostrukturanzeiger; M_1 = Teilmenge der Trägermenge, deren Elemente alle adressierten Angaben sowie alle bezugsadressentragende Angaben und damit alle Adressen sind

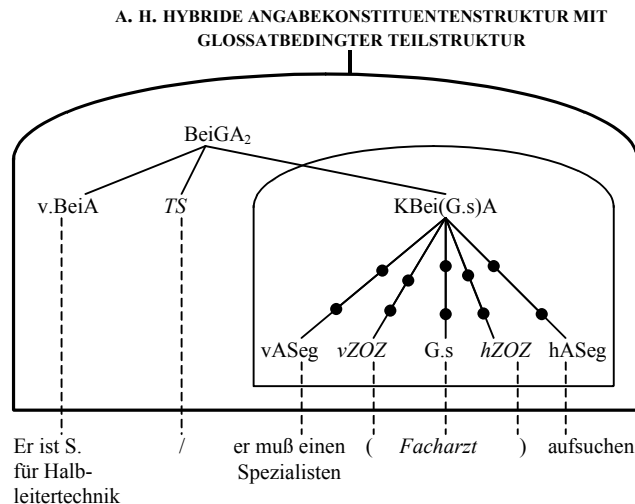


Abb. 16: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen hybriden Angabekonstituentenstruktur mit glossatbedingter Teilstruktur, die die Beispielgruppenangabe a_5 aus wa_5 in Abb. 1 aufweist (vgl. Abb. 12).

hybriden Kommentarmikrostrukturen. Ordnet man sie aber in die Typologie der hierarchischen hybriden Angabestrukturen ein, in der der Kommentarstatus von Angaben keine Rolle spielt, gehört sie in ihrer konkreten und abstrakten Ausprägung zum Typ der hierarchischen internlemmatisch architektonisch angereicherten Angabemikrostruktur (vgl. Abb. 20). Entsprechendes gilt für die zugehörigen konkreten Strukturen. — Die in Abb. 12 dargestellte abstrakte hierarchische hybride rechte Kernstruktur gehört als Teilstruktur der abstrakten hierarchischen hybriden Artikelstruktur zu den abstrakten hierarchischen hybriden Kommentarmikrostrukturen. Innerhalb der Typologie der hierarchischen hybriden Angabestrukturen gehört sie aber in ihrer konkreten und abstrakten Ausprägung zum Typ der hierarchischen hybriden Angabemikrostruktur mit minimierter glossatbedingter Teilstruktur (vgl. Abb. 20). Häufig gehören hierarchische hybride Angabekonstituenten- und Angabemikrostrukturen in der Typologie der Kommentarstrukturen zu den hierarchischen hybriden Kommentarkonstituenten — bzw. zu den Kommentarmikrostrukturen. Es gibt aber auch viele Fälle, in denen nichtelementare Angaben ohne Kommentarstatus hierarchische hybride Angabekonstituentenstrukturen und als deren Teilstrukturen hierarchische hybride Angabemikrostrukturen aufweisen. Ein Beispiel findet sich in wa_5 in Abb. 1. Die Beispielgruppenangabe ($BeiGA_2$) „er ist S. für Halbleitertechnik; er muss einen Spezialisten (Facharzt) aufsuchen“

aufsuchen“ (= $a_{5i} a_5 \in \text{BeiGA}_2$) weist eine hierarchische hybride Angabekontstituentenstruktur mit glossatbedingter Teilstruktur auf und als deren Teilstruktur eine hierarchische hybride Angabemikrostruktur mit minimierter glossatbedingter Teilstruktur (vgl. Abb. 20). Die zuletzt genannte Angabekontstituentenstruktur ist in Abb. 16 dargestellt.

Auf der Ebene der Strukturdarstellung ist die abstrakte und isomorphe konkrete hierarchische hybride Angabemikrostruktur mit minimierter glossatbedingter Teilstruktur von a_5 dadurch erhältlich, dass man *erstens* in Abb. 16 die Kanten mit den Endknoten *TS*, *vZOZ* und *hZOZ* und *zweitens* die zugehörigen Repräsentationen der Element-Klassen-Beziehung (x---Y) sowie die zugehörigen Elemente in der konkreten Struktur tilgt.

wa₈: *bóhrent* = stechen, daß es ein Loch gibt; drehend stechen. Davon der *Bóhrer* = Bohrwerkzeug, Werkzeug zum Verbohren, wer bohrt. Zusammensf.: das *Bóhrmehl* = mehlig beim Bohren abfallende Holzfeilschén.

wa₉: der/ **Méter** {des Meters, *ritkán* die Meter, das ter, m}
Der Schrank ist 2 **Meter** {Sg} hoch und ein(en) **Meter** breit. *A szekrény 2 méter magas és egy méter széles.* Meine kleine Schwester ist ein(en) **Meter** fünfzig groß. *A húgom egy méter ötven magas.* Der Maurer baut eine Mauer von 30 **Meter** {Sg} Länge. *A kőműves egy 30 méter hosszú falat épít.*

wa₁₀: **klein** [klain] <Adj.> **1.** <Steig. reg.> 'in räumlicher Ausdehnung unter einem bestimmten (mittleren) Wert liegend'; ANT groß (1,2.1): *ein ~er Mann, Baum; ein ~es Zimmer, Haus, Format, Paket; ~e Hände: das Grundstück ist mir zu ~; der Anzug ist, die Schuhe sind mir zu ~ geworden; er ist ~er als ich; ein Wort ~* ('mit kleinen Buchstaben') *schreiben; /in den kommunikativen Wendungen/ ~, aber oho /wird von jmdm. gesagt, der nicht groß, aber bemerkenswert energisch, leistungsfähig ist/; ~, aber mein* ('nicht sehr groß, aber es gehört mir') /sagt jmd., wenn er betonen will, dass er mit Räumlichkeiten, Fahrzeugen od. Gegenständen zufrieden ist, weil sie ihm gehören, auch wenn sie klein sind/ **2.** <o. Steig.; nicht bei Vb.> 'einen relativ kurzen Zeitraum umfassend' /beschränkt verbindbar/: *eine ~e* (ANT große) *Pause; eine ~e Weile warten; einen ~en Moment, bitte!* **3.** <Steig. reg.; nicht bei Vb.> 'aus einer relativ geringen Anzahl, Menge bestehend'; ANT groß (4) /vorw. auf Gruppen, Mengen bez./: *eine ~e Familie, Gruppe, Herde; etw. im ~en Kreis besprechen; eine ~e Summe, Anzahl: eine ~e Summe Geld, Anzahl, Personen; ein ~es* (SYN 'niedriges 2') *Gehalt;*

wa₁₁: *Citadelle, f. –n: eine
bei einer Stadt gelegene,
sie beherrschende Festung.

Abb. 17: Wörterbuchartikel wa₈ aus Weigand 1873 (gekürzt), wa₉ aus Hollós 2001 wa₁₀ aus DGWDaF und wa₁₁ aus Sanders-WDS

In wa₈ ist „Zusammens.: das B ó h r m ë h l“ eine um „das“, eine Artikelangabe, binnerweiterter Angabe zur Komposition. Die letzte Teilangabe ist die zweifach oben erweiterte nichtlemmatische Kompositumangabe „ B ó h r m ë h l“. Der erste oben erweiternde funktionale Angabezusatz ist eine Wortakzentkennzeichnung und an das Angabeformsegment „o“ hinabadressiert. Der zweite oben erweiternde funktionale Angabezusatz ist eine Vokalidentifizierungskennzeichnung, mit der nach der „Vorrede“ in Weigand (1873, VII) das mit dem Angabeformsegmente genannte e „nach Jacob Grimms Vorgang im Altdeutschen“ als so genanntes „tiefes e“ identifiziert wird. Die abstrakte (und isomorphe konkrete) externlemmatisch architektonisch angereicherte Angabemikrostrukturen sind in Abb. 18 dargestellt.

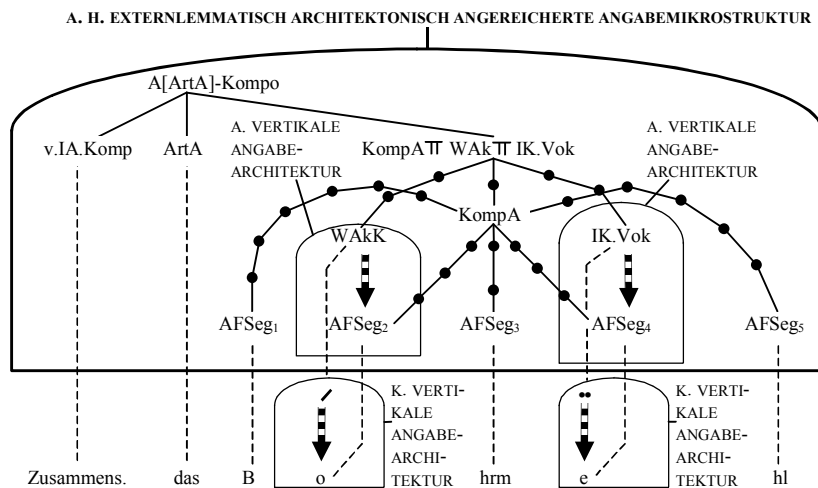


Abb. 18: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur (abstrakten und isomorphen konkreten) hierarchischen externlemmatisch architektonisch angereicherten Angabemikrostruktur der nichtlemmatischen um eine Artikelangabe (ArtA) binnerweiterter Angabe zur Komposition aus wa₈ (die Teilstrukturen der Artikelmikrostrukturen von wa₈ sind)

Abkürzungen: A[ArtA]-Kompo = um eine Artikelangabe binnerweiterter Angabe zur Komposition; v.IA.Komp = verdichtete Kompositumidentifizierungsangabe; KompA = Kompositumangabe; WakK = Wortakzentkennzeichnung; IK.Vok = Vokalidentifizierungskennzeichnung; „u —●— v“ bedeutet soviel wie u ist ein Teil von v, mit „u“ als Variable für Textsegmente, die keine Textkonstituenten sind; „a —●—> b“ bedeutet soviel wie a ist oberhalb von b, wenn der Pfeil nach unten zeigt; „┌“ bedeutet soviel wie oben erweitert um

Im Folgenden betrachten wir den bilingualen Wörterbuchartikel wa_9 , und zwar die Angabe eines äquivalenten Satzpaars „Der Maurer baut eine Mauer von 30 **Meter** {Sg} Länge. *A köműves egy 30 méter hosszú falat épít*“ (= a_9 ; $a_9 \in A.\ddot{a}SP$). Die nichtelementare Angabe besteht aus einer deutschen Kompetenzbeispielangabe, auf die als zweite zielsprachliche Teilangabe eine Angabe eines äquivalenten Satzes (A.äS) folgt. Die deutsche Kompetenzbeispielangabe ist durch einen unten erweiternden funktionalen Angabezusatz, bei dem es sich um eine durch einen Unterstrich realisierte Fehlerquellenkennzeichnung (von) handelt erweitert, sowie weiterhin durch ein grammatisches Binnenglossat „Sg“, bei dem es sich um eine Numeruskennzeichnung handelt. Damit weist die Angabe a_9 eine hierarchische externlemmatisch architektonisch angereicherte Angabekonstituentenstruktur mit glossatbedingter Teilstruktur auf, zu der als Teilstruktur eine hierarchische externlemmatisch architektonisch angereicherte Angabemikrostruktur mit minimierter glossatbedingter Teilstruktur gehört. Letztere ist in Abb. 19 dargestellt.

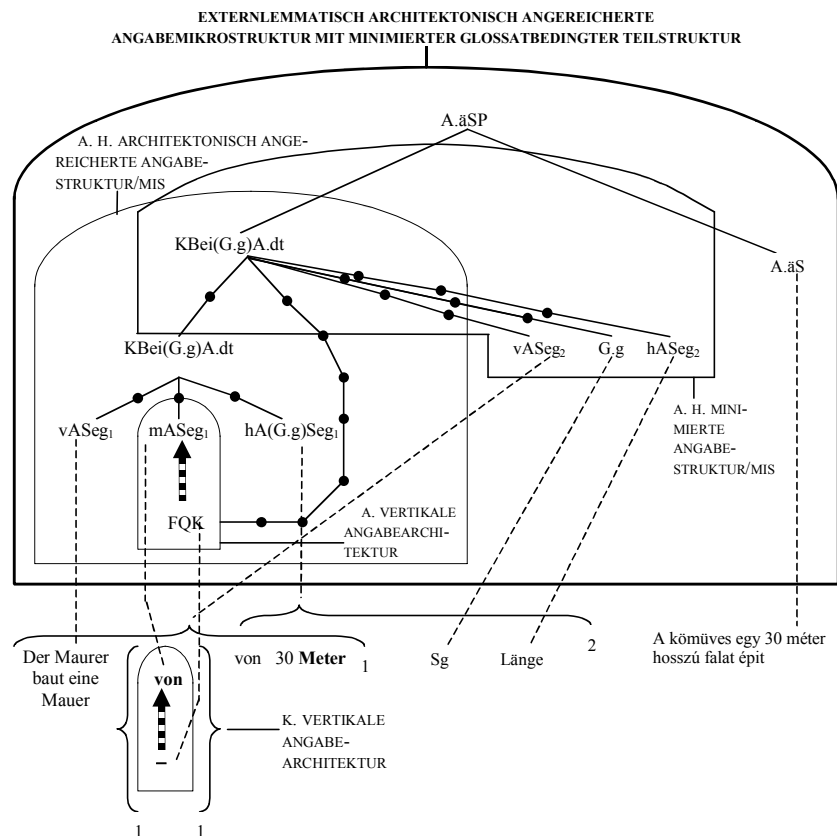


Abb. 19: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten (und konkreten isomor-

phen) hierarchischen externlemmatisch architektonisch angereicherten Angabemikrostruktur mit minimierter glossatbedingter Teilstruktur, die a_9 , die Angabe eines äquivalenten Satzpaares aus w_9 , aufweist

Abkürzungen: A.äSP = Angabe eines äquivalenten Satzpaares; A.äS = Angabe eines äquivalenten Satzes; KBei(G.g)A.dt = deutsche (dt.) um ein grammatisches Binnenglossat (G.g) erweiterte Kompetenzbeispielangabe; FQK = Fehlerquellenkennzeichnung „ $x \dashrightarrow y$ “ bedeutet soviel wie *x ist unterhalb von y*, wenn der Pfeil nach oben zeigt; *Darstellungskonventionen:* die mit 1 und 2 indizierten Klammern sollen helfen, die terminalen Elemente der konkreten Struktur den richtigen Klassen der abstrakten Struktur zuzuordnen

Die Typologie der hierarchischen hybriden Angabestrukturen ist reichhaltig. Einen kleinen Ausschnitt findet sich in Abb. 20.

Das mit dem Typologiegraph in Abb. 20 nur ein kleiner Ausschnitt aus einer Typologie für hierarchische hybride Angabestrukturen erfasst wird, wird sofort deutlich, wenn man bedenkt, dass hierarchische architektonisch angereicherte Angabekonstituenten- und Angabemikrostrukturen zusätzlich eine hierarchische hybride binnenerweiterungsbedingte bzw. eine hierarchische hybride minimierte binnenerweiterungsbedingte Teilstruktur aufweisen können, wie z.B. w_9 in Abb. 17 (vgl. Abb. 20).

6. Hybride artikelinterne Textkonstituentenstrukturen III: Angabestrukturen und exhaustive Angabestrukturen

Hierarchische Angabestrukturen sind — wie wir im Abschnitt 5 gesehen haben — Strukturen von erweiterten elementaren oder nichtelementaren Angaben. Hierarchische Angabestrukturen sind dagegen textuelle Strukturen von ganzen Wörterbuchartikeln und seltener solche von anderen akzessiven Einträgen, die hier nicht berücksichtigt werden. Ein hierarchische hybride Angabestruktur eines Wörterbuchartikels ist die um die nichthierarchische Adressierungsstruktur des gleichen Wörterbuchartikels erweiterte hierarchische hybride Artikelmikrostruktur; deswegen wird statt des Terminus *Angabestruktur* auch der synonyme Terminus *Mikro- mit Adressierungsstruktur* verwendet. Eine hierarchische hybride exhaustive Angabestruktur eines Wörterbuchartikels ist die um die nichthierarchische Adressierungsstruktur des gleichen Wörterbuchartikels erweiterte hierarchische hybride Artikelkonstituentenstruktur. Die Kenntnis dessen, was unter *Adressierung*, unter *Angabeadressierung*, unter *Angabezusatzadressierung* und besonders unter einer *artikelzugehörigen Adressierungsstruktur* verstanden wird, muss hier vorausgesetzt werden (vgl. z.B. Wiegand 2006 und WLWF-1 2010).

Wir betrachten zunächst hierarchische hybride Angabestrukturen. Verfügt man bereits über die konkrete und isomorphe abstrakte hierarchische hybride Artikelmikrostruktur sowie über die konkrete und isomorphe abstrakte Adressierungsstruktur eines Wörterbuchartikels, dann sind die artikel-

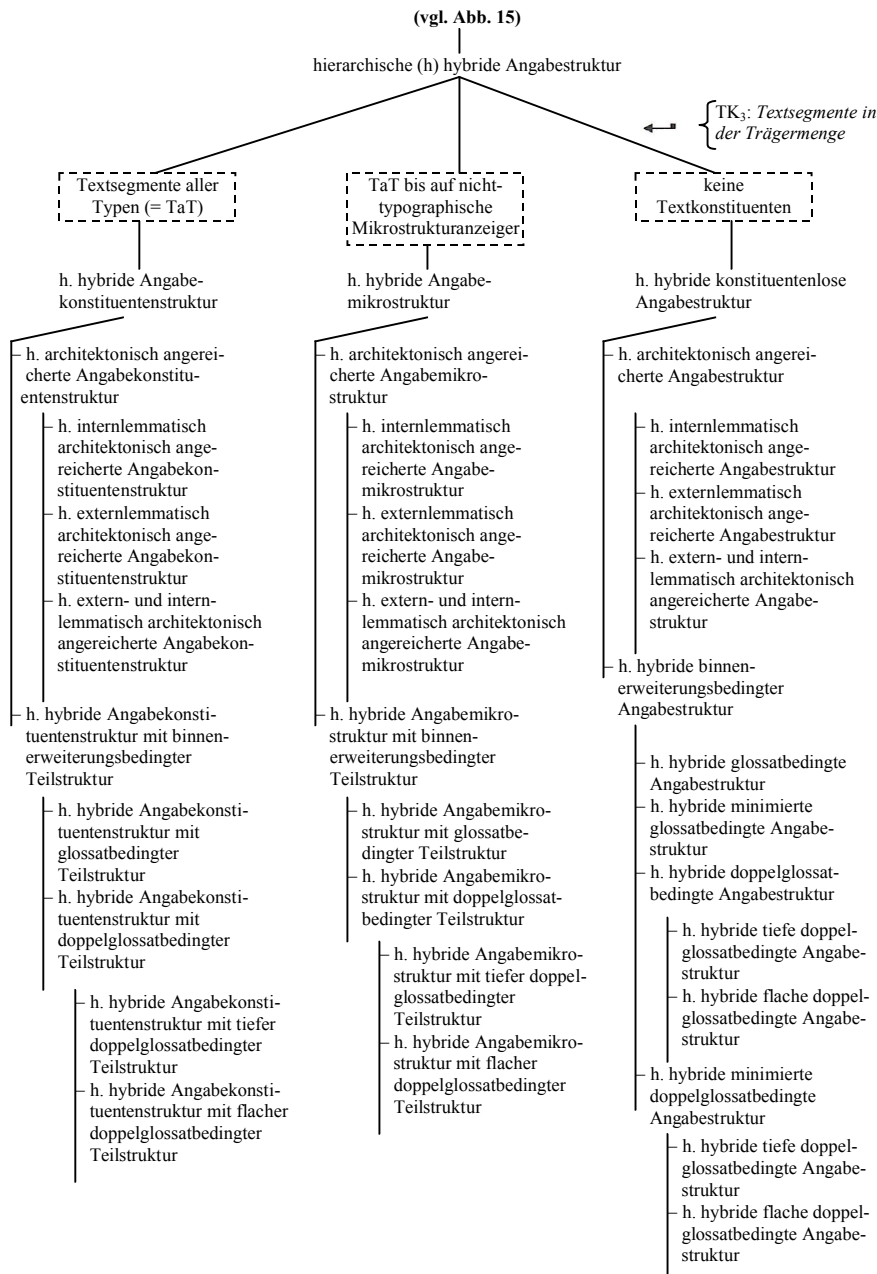


Abb. 20: Partiiell erweitert kommentierter Typologiegaph zu einem Ausschnitt aus einer Typologie von hierarchischen hybriden Angabestrukturen
Abkürzungen: h = hierarchisch; TK = Typologiekriterium; Darstellungskventionen: „←“ bedeutet soviel wie die Anwendung des TK führt zu der Unterteilung

zugehörigen konkrete und abstrakte Angabenstruktur dadurch erhältlich, dass man die beiden erstgenannten textuellen Strukturen zusammenfügt. Eine solche Strukturzusammenfügung lässt sich sowohl strikt formal als auch auf der Darstellungsebene vornehmen, wie es nachfolgende geschieht. Will man von der Darstellung einer hierarchischen hybriden Artikelmikrostruktur zur Darstellung einer hierarchischen hybriden Angabenstruktur übergehen, hat man mindestens die folgenden vier Darstellungsmethoden (vgl. Wiegand 2010):

- (a) Man repräsentiert innerhalb des Strukturgraphen jede Adressierungsbeziehung (analog zu den Pfeildiagrammen bei der Darstellung von Adressierungsstrukturen) durch einen Pfeil, der vom adressierten Textsegment zu dessen Bezugsadresse führt.
- (b) Man verwendet Adressensymbole, die die Bezugsadressen bezeichnen und die — getrennt durch einen fetten Mittenpunkt — hinter jedes Klassensymbol für eine Klasse von adressierten Textsegmenten gesetzt werden, das im Mikrostrukturgraphen als Knotenetikett fungiert. Ein entsprechendes erweitertes Knotenetikett hat z.B. die Form „MorA.S • LZGA“ und ist — vereinfacht ausgedrückt — zu lesen wie *an die LZGA adressierte MorA.S.*
- (c) Man verwendet Adressensymbole im Zusammenhang mit der Darstellungskonvention, dass bei adressenhomogenen Angaben die Verwendung der Adressensymbole bei allen Teilangaben der adressenhomogenen Mutterangabe unterbleibt, weil die Adressierungsbeziehungen erschließbar sind.
- (d) Man repräsentiert innerhalb des Mikrostrukturgraphen Adressierungsbeziehungen durch Pfeile im Zusammenhang mit der Darstellungskonvention, dass bei adressenhomogenen Angaben nur bei der Mutterangabe ein Pfeil zur Bezugsadresse führt, aber keiner von den adressengleichen Teilangaben der Mutterangabe, weil diese erschließbar sind.

Für die Strukturdarstellung von hierarchischen reinen und hybriden Angabenstrukturen kann die Darstellungskonvention in Geltung gesetzt werden, dass die Angabeadressierungsbeziehungen in der konkreten Struktur nicht explizit berücksichtigt werden, da sie aus der Darstellung in der abstrakten Struktur erschlossen werden können.

Die anschauliche Darstellungsmethode (a) eignet sich besonders zur Darstellung der hierarchischen reinen und hybriden Angabenstrukturen von Kurzartikeln sowie für die Darstellung solcher Kommentarangabenstrukturen (also Angabenteilstrukturen von artikelzugehörigen Angabenstrukturen), die zu einem Formkommentar gehören. In Abb. 21 findet sich die Darstellung der hierarchischen hybriden Kommentarangabenstruktur, die der Formkommentar von w_5 (= fk_5) in Abb. 1 aufweist nach der Darstellungsmethode (a) (vgl. Abb. 7).

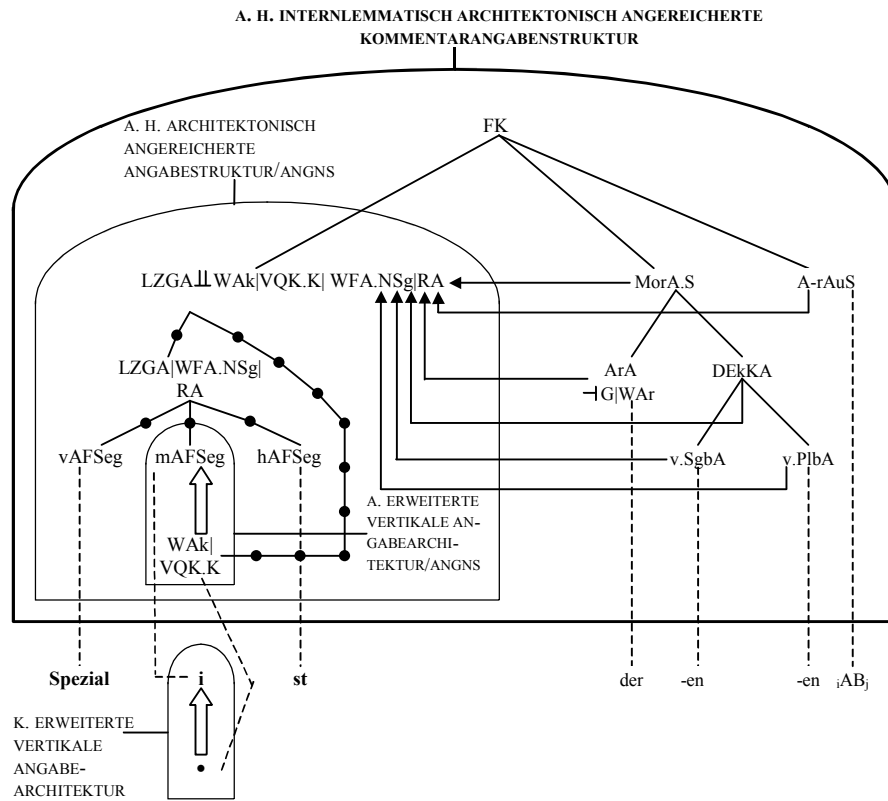


Abb. 21: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur konkreten (und isomorphen abstrakten) internlemmatisch architektonisch angereicherten Kommentarangabenstruktur, die fk_5 , der Formkommentar von wa_5 in Abb. 1, aufweist, „ $x \Rightarrow y$ “ bedeutet soviel wie x ist unterhalb von y und an y hinaufadressiert, wenn der Pfeil nach oben zeigt.

Betrachtet man die in Abb. 21 dargestellte hierarchische Struktur unter dem Aspekt, dass sie eine Teilstruktur der artikelzugehörigen hierarchischen hybriden Angabenstruktur ist, wird sie auch *hierarchische hybride linke Kernstruktur/AngnS* genannt. Dass in Abb. 21 *erweiterte vertikale Angabearchitekturen* vorliegen, liegt daran, dass auf der Trägermenge der Angabearchitektur nicht nur eine *unterhalb*-Relation, sondern auch eine zweistellige asymmetrische und irreflexive Relation vom Typ der Hinaufadressierungsrelation definiert ist. Denn die Wortakzentkennzeichnung, die zugleich eine Vokalquantitätskennzeichnung zur Kürze und durch den Unterpunkt realisiert ist, ist an das Angabeformsegment i hinaufadressiert.

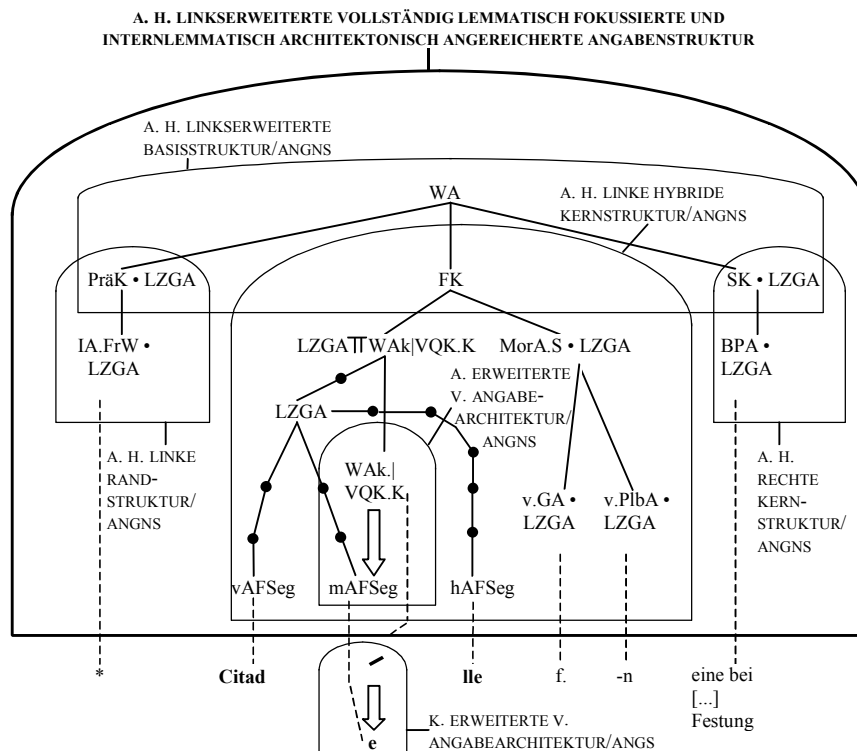


Abb. 22: Einfach kommentierter und mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen linkerweiterten vollständig lemmatisch fokussierten und internlemmatisch architektonisch angereicherten Angabenstruktur, die wa₁₁ in Abb. 17 aufweist

Abkürzungen: PräK = Präkommentar; IA.FrW = Fremdwortidentifizierungsangabe (realisiert durch einen Asterisk); v.GA = verdichtete Genusangabe; *Darstellungskonventionen:* „x ⇐⇒ y“ bedeutet soviel wie x ist oberhalb von y und an y hinabadressiert, wenn der Pfeil nach unten zeigt „u —●— v“ bedeutet soviel wie u ist ein Teil von v, mit „u“ als Variable für Textsegmente, die keine Textkonstituenten sind.

Verfügt man nicht über die hierarchische hybride Artikel mikrostruktur eines Artikels sowie zugleich über die artikelzugehörige nichthierarchische Adressierungsstruktur und möchte die hierarchische hybride Angabenstruktur eines Wörterbuchartikels wa_x bilden, dann muss man nach der Bildung der elementenheterogenen Trägermenge, die die gleichen Elemente aufweist als wolle man die hierarchische hybride Artikel mikrostruktur zu wa_x bilden, auf dieser Trägermenge zwei Relationen der uns bekannten Typen definieren, nämlich eine Relation vom Typ der partitiven und eine vom Typ der Präzedenzrelation,

sowie weiterhin auf den geeigneten Teilmengen der Trägermenge eine zweistellige, asymmetrische und irreflexive Relation vom Typ der Angabeadressierungsrelation mit dem Relationsterm *x ist adressiert an y* (mit „*x*“ als Variable für adressierte Angaben und „*y*“ als Variable für bezugsadressentragende Angaben und damit für Bezugsadressen) sowie schließlich eine zweistellige, asymmetrische und irreflexive Relation vom Typ der Angabezusatzadressierungsrelation mit dem Relationsterm *u ist adressiert an v* (mit „*u*“ als Variable für funktionale Angabezusätze und „*v*“ als Variable für die Träger von Angabezusatzadressen und damit für Angabezusatzadressen). Das Ergebnis ist dann eine hierarchische hybride Angabenstruktur wie die in Abb. 22 dargestellte von wa₁₁.

Im Folgenden betrachten wir ein Beispiel für hierarchische hybride exhaustive Angabenstrukturen eines Wörterbuchartikels anhand von wa₇ in Abb. 1 und auf der Basis von Abb. 14; das bedeutet, dass (aus Platzgründen) der Übergang von der Artikelkonstituenten — zur artikelzugehörigen Angabenkonstituentenstruktur auch hier lediglich auf der Darstellungsebene vorgenommen wird. Angewandt wird die oben erläuterte Darstellungsmethode (b), so dass die artikelinternen Angabeadressierungsbeziehungen dadurch repräsentiert wird, dass hinter jedes Klassensymbol für eine Klasse von adressierten Textsegmenten hinter einem fetten Mittenpunkt das Adressensymbol steht. Während in wa₁₁ in Abb. 17 die Adressierungskonstellation der vollständig lemmatischen Adressierung mit internlemmatischer Adressierung (vgl. Wiegand 2006: 230ff u. WLWF-1 2010) vorliegt (jede internlemmatische Adressierung ist eine Angabezusatzadressierung), so dass alle adressierungsfähigen Angaben an die als lemmatische Substantivangabe ausgeprägte Lemmzeichengestaltangabe adressiert sind, und deswegen nur das Adressensymbol LZGA vorkommt (vgl. Abb. 22), liegt in wa₆ in Abb. 1 die Adressierungskonstellation der partiell lemmatischen Adressierung mit externlemmatischer Angabezusatzadressierung vor. Das bedeutet auch, dass mehrere verschiedene Adressensymbole auftreten, denn im Postkommentar zu Phraseologie sind nicht alle Angaben an die Lemmzeichengestaltangabe adressiert, und auch die Glossate sind nichtlemmatisch adressiert. Es ist klar, dass die Klassensymbole für die nichttypographischen Mikrostrukturanzeiger (z.B. *TS*) nicht mit einem Adressensymbol versehen werden, da nichttypographische Mikrostrukturanzeiger nicht adressiert sind. Ein Ausschnitt aus der abstrakten (und isomorphen konkreten) partiell lemmatisch fokussierten hybriden exhaustiven Angabenstruktur mit tiefer doppelglossatbedingter Teilstruktur findet sich in Abb. 23.

Im Folgenden betrachten wir die in Abb. 23a nur markierte und nicht analysierte hybride tiefe glossatbedingte Angabe- mit Adressierungsstruktur, die die verdichtete und um ein semantisches Doppelglossat binnenerweiterte Kompetenzbeispielangabe „das hat viel S (*Mühe, Anstrengung*) gekostet“ (= a₆) in wa₆ in Abb. 1 aufweist und die in Abb. 23b dargestellt ist.

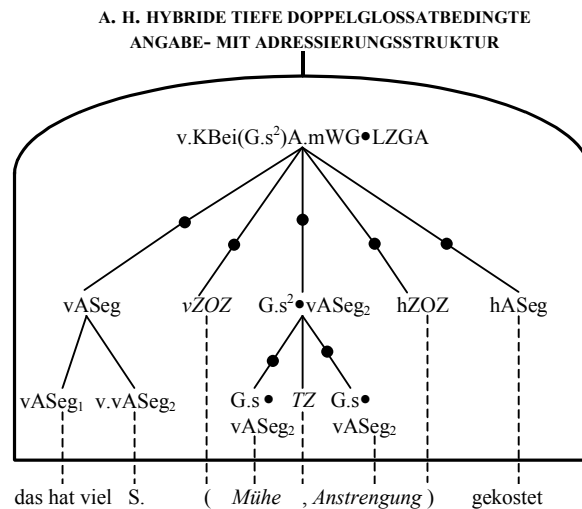


Abb. 23b: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) tiefen doppelglossatbedingten Angabe- mit Adressierungsstruktur
Abkürzungen: v.vASeg = verdichtetes vorderes Angabesegment

Oben wurde festgestellt, dass Angabenstrukturen textuelle Strukturen von ganzen Wörterbuchartikeln sind. Es gibt aber auch Angaben, wie z.B. a_6 , innerhalb derer Adressierungsbeziehungen auftreten; so ist in a_6 das erste semantische Binnenglossat „Mühe“ adjazent links an „S.“ adressiert, und das zweite semantische Binnenglossat ist nichtadjazent links an „S.“ linksadressiert. Um nicht einen zu leicht missverständlichen Terminus wie *angabeinterne Angabenstruktur* verwenden zu müssen, wird daher von *Angabemikrostruktur mit Adressierungsstruktur* oder verkürzt von *Angabe- mit Adressierungsstruktur* gesprochen (in Analogie zu dem oben erwähnten terminologischen Synonym *Mikro- mit Adressierungsstruktur* zu *Angabenstruktur*).

Um die in Abb. 23b dargestellte Angabe- mit Adressierungsstruktur, die zu a_6 gehört, bilden zu können, muss das vordere Angabesegment „das hat viel S.“ weiter nichtfunktional-positional segmentiert werden, damit als Teilsegment des verdichteten vorderen Angabeformsegments das „S.“ isoliert wird, das Adressenträger der Bezugsadressen $S|$ der beiden semantischen Binnenglossate ist. Erst dann kann auf den zuständigen Trägermengen für die konkrete und abstrakte Strukturausprägung neben der partitiven und der Präzedenzrelation eine zweistellige asymmetrische und irreflexiven Relation vom Typ der Angabezusatzadressierungsrelation definiert werden, so dass die in Abb. 23b dargestellte hybride tiefe doppelglossatbedingte Angabe- mit Adressierungsstruktur erhältlich ist.

Abschließend sei zu den artikelinternen hierarchischen reinen und hybriden Angabenstrukturen und exhaustiven Angabenstrukturen bemerkt, dass sie für theoriegeleitetes Verfassen von Benutzungshinweisen die am besten geeigneten artikelzugehörigen textuellen Strukturen sind, weil sie — anders als die hierarchischen reinen und hybriden Artikelkonstituenten- und Artikelmikrostrukturen — Auskunft darüber geben, auf welche Angabegegenstände anderer artikelinterner Angaben die adressierten Angaben und adressierten funktionalen Angabezusätze zu beziehen sind, wenn ein Benutzer anhand artikelinterner Angaben und funktionaler Angabezusätze eine lexikographische Information (verstanden als punktuelles Wissen und damit als kognitives Phänomen) erschließen möchte (vgl. dazu Wiegand 2000b).

7. Hybride textuelle Einheiten: eine kleine Auswahl

Wir wissen bereits, dass die funktional-positionale Segmentation von Textsegmenten, die selbst das Ergebnis einer nichtfunktionalen Segmentation und damit keine Textkonstituenten sind, wie das (nichtelementare) semantische Doppelglossat „*Mühe, Anstrengung*“, hybride Textkonstituenten sind, da sie *von zweierlei Herkunft* sind (vgl. die Bedeutungsparaphrasenangabe des Duden⁴GFWb in 1.).

Potenziell nichtelementare Glossate jeder Art und nicht nur Doppelglossate haben, wenn sie im Kotext eines Wörterbuchartikels oder in dem eines der anderen akzessiven Einträge auftreten, die Eigenschaft aller Binnenglossate: Sie sind funktional-positional nicht isolierbar, weil bei der Segmentation der Angabe, die sie binnenerweitern, nichtfunktionale Textsegmente, nämlich vordere und hintere Angabesegmente auftreten. Damit sind sie funktionale Angabezusätze ohne Textkonstituentenstatus und weisen die kriteriale Eigenschaft der Elemente dieser Textsegmentklasse auf. Sind sie aber nichtfunktional-positional isoliert, dann weisen sie kriteriale Eigenschaft der Textkonstituenten auf: Sie sind dann aktuell nichtelementar und damit restfrei in funktionale Textsegmente durch eine Anwendung der Methode der funktional-positionalen Segmentation zerlegbar.

Dies sei im Folgenden anhand zweier anderer Typen von potenziell nichtelementaren Binnenglossaten gezeigt (vgl. Wiegand 2010a). Im Artikel wa_{10} in Abb. 17 findet sich folgende elementare Angabe: „*eine ~e (ANT große) Pause*“ (= a_{10}). Es handelt sich um eine verdichtete Kompetenzbeispielangabe, die um ein semantisches Binnenglossat erweitert ist, bei dem es sich um eine Kennzeichnung zur Antonymie handelt, die aus „ANT“, einer verdichteten Antonymidentifizierungskennzeichnung, sowie aus „große“, einem Antonymhinweis besteht. Um die konkrete hierarchische glossatbedingte Angabestruktur von a_{10} bilden zu können, muss a_{10} nichtfunktional-positional segmentiert werden, und zwar wie folgt: $eine \sim e \mid (\mid ANT \text{ große } \mid) \mid Pause$ (mit „ \mid “ als Segmentationsfugenmarkierung). Es gelten folgende Aussagen: $eine \sim e \in vASeg$; $(\mid ANT \text{ große } \mid) \in hZOZ$ und $Pause \in gASeg$. Alle fünf Textseg-

mente sind keine Textkonstituenten. Das mittlere Angabesegment (mASeg), die isolierte Kennzeichnung zu Antonymie, lässt sich restfrei funktional-positional in zwei Angaben segmentieren und ist damit — wie das Doppelglossat aus wa₆ — als hybride Textkonstituente einzustufen. Die abstrakte (und isomorphe konkrete) hierarchische tiefe glossatbedingte Angabestruktur findet sich in Abb. 24.

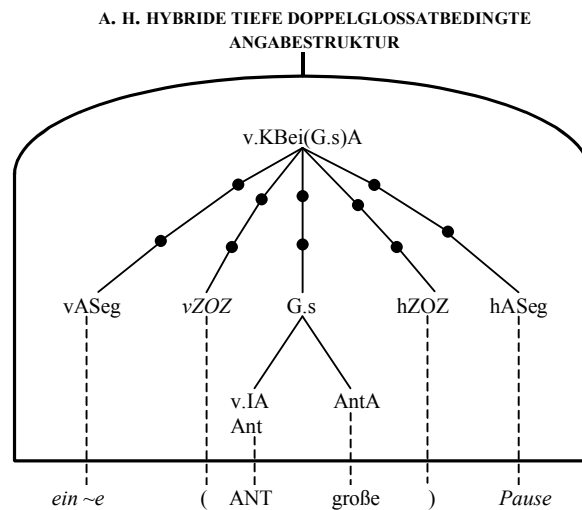


Abb. 24: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) hierarchischen hybriden glossatbedingten Angabestruktur, konzipiert als tiefe Struktur
Abkürzungen: v.IA.Ant = verdichtete Antonymenidentifizierungsangabe; AntA = Antonymangabe

Auch die um eine Kennzeichnung zur Synonymie binnerweiterter Kompetenzbeispielangabe „*ein ~es* („SYN 'niedriges' 2“) *Gehalt*“ aus wa₈ in Abb. 17, deren semantisches Binnenglossat aus einer verdichteten Synonymidentifizierungsangabe „SYN“ und aus einem um eine Bedeutungsidentifizierungskennzeichnung rechtserweiterten Synonymhinweis „'niedriges' 2“ besteht, lässt sich analog zum vorhergehenden Beispiel analysieren, so dass „Syn 'niedriges' 2“ als hybride Textkonstituente einzustufen ist.

Als letztes Beispiel sei anhand von wa₉ und wa₁₀ in Abb. 25 erläutert, was unter einer hybriden Verweiskennzeichnung zu verstehen ist.

wa₉: **Dorf**, das; -s, Dörfer.
+ die ↗ Kirche im D. lassen

wa₁₀: **Mineralsalz**, das; es, -e: ↑ anorganisches Salz, das sowohl in der Natur vorkommt als auch künstlich hergestellt wird.

Abb. 25: Wörterbuchartikel wa₉ (gekürzt) aus HWDG und wa₁₀ (gekürzt) aus Duden-4GFWb

Funktionale Angabezusätze vom Typ der Verweiskennzeichnung, wie z.B. „↗ Kirche“, in der verdichteten verweisvermittelnden Phrasemangabe in wa₉ in Abb. 25, die nichthybrid sind, weisen eine Verweisbeziehungs-kennzeichnung auf — in wa₉ realisiert durch den Schrägpfel — die im Unterschied zu erweiternden Verweisbeziehungsangaben, wie „↑“ in wa₁₀, funktional-positional nicht isolierbar sind. In wa₁₀ ist die Definiensangabe um den Hochpfeil „↑“ linkerseitig erweitert. Durch diese Linkserweiterung ergibt sich die usuelle Verweiskennzeichnung „↑ anorganisches Salz“; diese ist als ganzes Textsegment nicht funktional-positional isolierbar. Bei einer funktional-positionalen Segmentierung von wa₁₀ ist der Hochpfeil jedoch eine eigenständige Textkonstituente, mit der eine erweiternde Verweisbeziehungsangabe gegeben ist, und zwar ein linkerseitig erweiternde. Dies bedeutet: Verweiskennzeichnungen mit linkerseitig erweiternden Verweisbeziehungsangaben sind hybride funktionale Angabezusätze, die keine Textkonstituenten sind.

8. Schlussbemerkung

Wörterbuchartikeln, die mindestens einen funktionalen Angabezusatz aufweisen, können — wie gezeigt wurde — reine und hybride textuelle Strukturen vom Typ der Artikelkonstituenten- und vom Typ der Artikelmikrostruktur, weiterhin solche vom Typ der artikelinternen Angabenstruktur und schließlich solche vom Typ der artikelinternen exhaustiven Angabenstruktur zugewiesen werden. Entsprechendes gilt für erweiterte Angaben und für nichtelementare um funktionale Angabezusätze erweiterte Angaben. In allen Fällen sind die hybriden textuellen Strukturen in einem höheren Grad explizit, und vor allen Dingen erlauben sie leichter eine genaue quantitative Analyse der propositionalen Gehalte, die der Lexikograph dem Benutzer vermitteln möchte. Dies gilt besonders für die hybriden Artikelmikrostrukturen, die als Distributionsstrukturen für alle funktionalen Textsegmente mit Angabefunktion im Wörterbuchartikel verstanden werden müssen. Um ein Wörterbuch mit hybriden Strukturen und Einheiten erfolgreich benutzen zu können, muss das Wissen des kundigen Benutzers (i.S.v. Wiegand 1998: 506) größer sein als wenn nur reine Strukturen gegeben sind.

Literatur

- Bergenholtz, Henning, Sven Tarp und Herbert Ernst Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L., H. Kalverkämper und H.E. Wiegand (Hrsg.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832*. Handbücher zur Sprach- und Kommunikationswissenschaft 14.2. Berlin/New York: Walter de Gruyter.
- Bourbaki, Nicolas. 1957. *Éléments de mathématique. I. Les structures fondamentales de l'analyse, livre I: Théorie des ensembles, chapitre 4: Structures*. Paris: Hermann [engl. Fassung 1968].

- DGWDaF 2000** = Kempcke, G. (Hrsg.). *Wörterbuch Deutsch als Fremdsprache*. Berlin/New York: Walter de Gruyter.
- Duden-⁴GFwB 2007** = *Das grosse Fremdwörterbuch. Herkunft und Bedeutung der Fremdwörter*. 4., aktual. Aufl. Hrsg. u. bearb. vom Wiss. Rat der Dudenredaktion. Mannheim [usw.]: Dudenverlag.
- Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand und Zgusta Ladislav (Hrsg.)**. 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. 3 Teilbände. Berlin/New York: Walter de Gruyter.
- Hollós, Zita**. 2001. *Német-Magyar suliszótár*. Szeged: Grimm Könyvkiadó.
- HWdG 1984** = *Handwörterbuch der deutschen Gegenwartssprache*. In zwei Bänden. Von einem Autorenkollektiv unter der Leitung v. G. Kempcke [...] *Bd. 1: A–K; Bd. 2: L–Z*. Berlin: Akademie-Verlag.
- Sanders-WDS 1969** = Daniel Sanders: *Wörterbuch der deutschen Sprache. Mit Belegen von Luther bis auf die Gegenwart*. 1. Bd. A–K. 2., unv. Abdruck. Leipzig: O. Wigand, 1876 [Nachdruck Hildesheim 1969. Mit einer Einführung und Bibliographie v. Werner Betz].
- Wiegand, Friedrich Ludwig Karl**. 1873. *Deutsches Wörterbuch*. 1. Bd. A–L. 2., verb.u. verm. Auflage. Gießen: J. Richter'sche Buchhandlung.
- Wiegand, Herbert Ernst**. 1989. Der Begriff der Mikrostruktur. Geschichte, Probleme, Perspektiven. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Hrsg.) 1989–1991: 409–462.
- Wiegand, Herbert Ernst**. 1989a. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Hrsg.) 1989–1991: 462–501.
- Wiegand, Herbert Ernst**. 1990. Printed Dictionaries and their Parts as Texts. An Overview on Recent Research as an Introduction. *Lexicographica* 6: 1–126 [Auch in Wiegand 2000: 951–1062].
- Wiegand, Herbert Ernst**. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Mit 159 Abbildungen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst**. 2000. Kammerer, Matthias und Werner Wolski (Hrsg.). *Kleine Schriften. Eine Auswahl aus den Jahren 1970 bis 1999 in zwei Bänden*. Bd. 1: 1970–1988; Bd. 2: 1988–1999. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst**. 2000a. Wissen, Wissensrepräsentation und Printwörterbücher. Heid, U., St. Evert, E. Lehmann und Ch. Rohrer (Hrsg.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8–12.2000*: 15–38. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Wiegand, Herbert Ernst**. 2001. Sprachkontaktwörterbücher, Typen, Funktionen, Strukturen. Iglá B., P. Petkov und H.E. Wiegand (Hrsg.). 2001. *Theoretische und praktische Probleme der Lexikographie*. 1. *Internationales Kolloquium zur Wörterbuchforschung am Institut Germanicum der St. Kliment Ohridski-Universität, Sofia, 7. bis 8. Juli 2000*: 115–224. Germanistische Linguistik 161–162: Hildesheim/Zürich/New York: Georg Olms Verlag.
- Wiegand, Herbert Ernst**. 2002. Über textuelle Strukturen der Wörterbuchartikel und Artikel-nischen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Zugleich ein Beitrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, H.E. (Hrsg.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuchs Deutsch als Fremdsprache*: 497–595. *Lexicographica. Series Maior* 110. Tübingen: Max Niemeyer Verlag.

- Wiegand, Herbert Ernst.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, Herbert Ernst.** 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.
- Wiegand, Herbert Ernst.** 2006. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 22: 187-261.
- Wiegand, Herbert Ernst.** 2009. Nichtnatürlich über natürliche Sprache schreiben. Zu einigen formalen Aspekten von Wörterbuchartikeln. *Heidelberger Jahrbücher* 53: 247-285.
- Wiegand, Herbert Ernst.** 2009a. Hybrid Text Constituent Structures of Dictionary Articles. A Contribution of the Expansion of the Theory of Textual Dictionary Structures. Nielsen, S. und S. Tarp (Hrsg.). 2009. *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*: 63-89. Terminology and Lexicography Research and Practice 12. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, Herbert Ernst.** 2010. Zur Methodologie der Wörterbuchforschung: Ausgewählte wörterbuchbezogene Analyse- und Darstellungsmethoden. *Lexicographica* 26. [Erscheint].
- Wiegand, Herbert Ernst.** 2010a. Semantik, Pragmatik und Wörterbuchform in einsprachigen Wörterbüchern. *Zeitschrift für germanistische Linguistik* 38(3). [Erscheint].
- Wiegand, Herbert Ernst und M^a Teresa Fuentes Morán.** 2009. *Estructuras lexicográficas. Aspectos centrales de una teoría de la forma del diccionario*. Granada: Tragalacanto.
- WLWF-1.** 2010 = *Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen/With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages*. Hrsg. u. Bearb. v./Ed. and compiled by Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer, Werner Wolski unter Mitarbeit von/with the collaboration of Ekaterina Butina-Koller (Russisch/Russian), Rute Costa (Portugiesisch/Portuguese), M^a Teresa Fuentes Morán (Spanisch/Spanish), Laura Giacomini (Italienisch/Italian), Rufus H. Gouws (Afrikaans; Englisch/English), Franz Josef Hausmann und Maria Hegner (Französisch/French), Regina Hessky und Zita Hollós (Ungarisch/Hungarian), Pavel Petkov (Bulgarisch/Bulgarian), Giovanni Rovere (Italienisch/Italian), Stefan J. Schierholz (Portugiesisch/Portuguese), Maria Smit (Englisch/English). 4 Bde./4 Vols. 1. Bd.: *Systematische Einführung/Vol. I: Systematic Introduction. A-C*. Mit 120 Grafiken von/with 120 Figures by Matthias Kammerer. Berlin/New York: Walter de Gruyter. [Mit/with CD-ROM].

Countable or Uncountable? That is the Question — Lexicographic Solutions to Nominal Countability in Learner's Dictionaries for Production Purposes

Mei Xue, *Centre for Lexicography, Aarhus School of Business, Aarhus University, Aarhus, Denmark (meix@asb.dk)*

Abstract: Nominal countability is problematic for foreign learners in production-oriented activities, especially for those whose native language does not have the corresponding function systems. In contrast to foreign learners' great difficulties with nominal numbers in foreign text production as shown in extensive linguistic research, there is a scarcity of lexicographic research on assisting foreign learners with their difficulties in this respect. Prompted by the great discrepancy between foreign learners' needs for lexicographic assistance and the relative indifference concerning this in lexicographic research, this study attempts to explore to what extent the lexicographic data in connection with nominal countability and their presentation in present learner's dictionaries can help foreign learners in foreign text production. This study will offer some constructive lexicographic solutions to the problems incurred by foreign learners' misconceptions of nominal countability in foreign text production.

Keywords: CHINESE EFL LEARNERS, CHINESE LEARNERS' DIFFICULTIES, DATA ACCESSIBILITY, DATA PRESENTATION, FOREIGN TEXT PRODUCTION, LEXICOGRAPHIC INFORMATION NEEDS, INFORMATION RETRIEVAL, LEARNER'S DICTIONARIES, LEXICOGRAPHIC SOLUTIONS, NOMINAL COUNTABILITY, PROFILE OF CHINESE LEARNERS

Opsomming: **Telbaar of ontelbaar? Dit is die vraag — Leksikografiese oplossings vir naamwoordelike telbaarheid in aanleerderswoordeboeke vir produksiedoeleindes.** Nominale telbaarheid is problematies vir vreemde aanleerders by produksiegerigte handelinge, veral vir hulle wie se moedertaal nie die ooreenstemmende funksiesisteme het nie. In teenstelling met vreemde aanleerders se groot moeilikhede met nominale getalle in vreemdeteksproduksie soos uit omvattende taalkundige navorsing blyk, is daar 'n skaarste aan leksikografiese navorsing wat vreemde aanleerders help met hul moeilikhede in hierdie opsig. Na aanleiding van die groot verskil tussen vreemde aanleerders se behoefte aan leksikografiese hulp en die relatiewe onbelangstelling hieroor in leksikografiese navorsing, probeer hierdie studie nagaan tot watter mate die leksikografiese data in verband met nominale telbaarheid en die aanbieding hiervan in huidige aanleerderswoordeboeke vreemde aanleerders in vreemdeteksproduksie kan help. Hierdie studie sal enkele konstruktiewe leksikografiese oplossings aanbied vir probleme ondervind as gevolg van vreemde aanleerders se wanopvatting oor nominale telbaarheid by vreemdeteksproduksie.

Sleutelwoorde: SJINESE EVT-AANLEERDERS, MOEILIKHEDE VAN SJINESE AANLEERDERS, DATATOEGANKLIKHEID, DATA-AANBIEDING, VREEMDETEKSPRODUKSIE, LEKSIKOGRAFIESE INLIGTINGSBEHOEFTE, INLIGTINGSHERWINNING, AANLEERDERSWOORDEBOEKE, LEKSIKOGRAFIESE OPLOSSINGS, NOMINALE TELBAARHEID, PROFIEL VAN SJINESE AANLEERDERS

1. Introduction

Attention to the countability of nouns in learner's dictionaries can be traced to as early as 1938 in Palmer's *Grammar of English Words*. Later, Hornby et al. used the abbreviations [C] and [U] to indicate the nominal countability in their *Idiomatic and Syntactic English Dictionary* (Cowie 1999: 46). Present learner's dictionaries follow the practice of using the labels [C] and [U] as well as the further specifications like singular and plural to indicate the numeral aspects of a noun. Miller (2006: 435) seems to be overconfident of such a practice by stating that "the inclusion of countability in learner's dictionaries has provided successive generations of students with a ready-made tool to help them acquire one of the hardest grammatical features of the English language". The inclusion of information related to nominal countability in learner's dictionaries will assist foreign learners in certain situations, but it is doubtful whether the mere act of including such data will automatically be helpful. Even Miller's (2006: 439) own study shows that there is a very subtle "difference in increased correctness between the dictionary and non-dictionary users", but she attributes the causes of her findings to learners' "lack of familiarity with the dictionaries, and their reluctance to check every noun in the dictionary for countability".

The assumption can seemingly be made that the present lexicographic practice of indicating information in connection with nominal countability in learner's dictionaries is transparent enough for foreign learners to confront their difficulties in this respect. However, it is doubtful whether the information retrieval process is as transparent as promised considering the fact that foreign learners with different language and cultural backgrounds may have a different conceptual understanding of the numeral aspects of a noun, and therefore have different lexicographic information needs. Despite the occasional doubt expressed about the present lexicographic practice, so far little serious attention has been paid to discussing the possible lexicographic solutions to foreign learners' difficulties in connection with nominal countability in foreign language learning.

Hornby (1965: 109-110) earlier identified foreign learners' difficulties regarding the numeral aspects of nouns in English, such as singular, plural, countable and uncountable. To address foreign learners' misuse of articles caused by their misconception of countability, he proposes that "the dictionary must show, by the provision of examples, when these phrases with significant article omission are possible and what the omission of the article indicates".

Later, Hausmann and Gorbahn (1989: 52) mention that contextual usage examples are needed for learners' correct application of the concept of countability of nouns, and point out that the mere marking of countability in the form of [C] and [U] without further explanations in some noun entries such as *hair* will confuse foreign learners engaged in production activities. But, as the aim of their study is to make a comprehensive comparison of two dictionaries (e.g. COBUILD and LDOCE II), the discussion about the lexicographic treatment of the countability of nouns is not pursued further. Discussing dictionary grammars, Andersen (2007: 121) proposes that the data related to nominal countability should be presented in a specific grammar part in view of the difficulty in indicating possible verb forms following collective nouns in an economical way or inventing economical codes to illustrate various determiners with nouns. However, this study is more concerned with the specific usage situation in which a noun is used, rather than the user situation giving rise to the varieties of usage situations in which a noun is involved when referring to the presentation of the data related to nominal countability.

In contrast to the scarcity of lexicographic research on nominal countability, extensive research in applied linguistics has revealed the evident difficulties foreign learners experience in production activities and the various types of errors incurred by the misconception of nominal countability, including the misuse of the English article system, the wrong use of quantifiers and premodifiers, the violation of subject-verb agreement, etc. (cf. Swan and Smith 1987; Sinclair 1991; Master 1997, 2002; Robertson 2000; Milton 2001; Chuang 2005). The asymmetry between foreign learners' practical difficulties caused by nominal countability in foreign language learning and the indifference shown in lexicographic research prompts this study to explore to what extent the present lexicographic practice of indicating nominal countability in learner's dictionaries can help foreign learners in foreign text production. It is followed by an attempt to propose some suggestions for future lexicographic practice. The profile of foreign learners as dictionary users shaped by their native language and social cultural contexts will be highlighted, as foreign language learning cannot take place in a vacuum.

This study consists of the following three stages. First, it identifies the semantic, syntactic as well as pragmatic difficulties in connection with nominal countability experienced by a specific group of learners in a specific user situation. It focuses on Chinese learners of English at college level in writing English under foreign language learning circumstances. Second, it examines the present lexicographic treatment of nominal countability in major learner's dictionaries in print as well as online formats. Some articles are analyzed to explore how to optimize appropriate lexicographic assistance with the countability concept sought by Chinese learners in English writing. Finally, suggestions are proposed and demonstrated in example articles for future lexicographic research and practice. As this study is limited to English learner's dictionaries, so reference to dictionaries in the following is to English learner's dictionaries only.

2. Chinese learners' lexicographic information needs for nominal countability in foreign text production

Foreign learners seldom look for information about nominal countability in dictionaries to comprehend foreign texts except when the ambiguity in comprehension is caused by the numeral features of a noun or the numeral inflection of a noun which may prevent them from identifying the canonical word form. Unlike the primary pursuit of meaning in foreign text reception, production-oriented activities demand fluency as well as accuracy, so the word form regarding nominal countability becomes evident. The numeral features of a noun will decide the determiners or quantifiers preceding a noun as well as the verb forms following it. The significance of appropriately using a noun in syntactic environments may be interpreted at lexical, semantic, grammatical and pragmatic level. The difficulties of foreign learners in this respect have been observed in many studies. Chinese learners of English engaged in English writing are no exception. Normally dictionaries, unlike grammar books, will have been observed to address such difficult problems, because learner's dictionaries are generally assumed to provide a variety of specific information about individual words while grammar books mainly focus on general grammatical rules.

In addition, in Chinese learning culture, dictionaries are regarded as almost unchallenged authorities of knowledge. So, when Chinese learners are uncertain of the countability features of a noun which they intend to use, they may seek assistance from available dictionaries to assure themselves of the appropriate syntactic behaviour of a noun to achieve accuracy in cross-cultural communication. As the overriding function of learner's dictionaries is to satisfy learners' punctual information needs regarding foreign language learning (Tarp 2008: 137), it is therefore essential to identify Chinese learners' information needs for lexicographic assistance regarding nominal countability, and to discuss how to present the data relevant to nominal countability to cater for Chinese learners' lexicographic needs in writing English.

2.1 The profile of Chinese learners

A description of the profile of Chinese learners will shed some light on understanding their particular lexicographic information needs concerning nominal countability, because their specific needs for lexicographic assistance is greatly subject to their profile in foreign language learning situations. The profile of Chinese learners in this study can be described by the following parameters: their native language, their learning contexts, the didactics, and their world knowledge. Although learners at college level are normally assumed to be at intermediate level in the Chinese educational context regarding their English proficiency, they still resort to their native language (e.g. generally Mandarin Chinese) when they are impeded in formulating their ideas. Their world

knowledge determined by their socio-cultural context tends to colour the concepts they attempt to express in their English writing.

To be more specific: This group of Chinese learners is normally at the age ranging from 17 to 20. Upon reaching college-level education, they have usually been learning English in formal classroom context for 6-9 years or more. The concept of nominal countability and the relevant grammar is assumed to be taught consciously in class in the first year of junior middle school when most Chinese learners start their formal English learning. Although nominal countability is clearly stated as one of the important teaching and examining points in the English teaching syllabus administered by the Chinese Ministry of Education, the interpretation and implementation of the teaching syllabus largely rest with individual English teachers. Therefore, when the uneven education resources distributed in the large geographic areas of China is considered, it is highly doubtful whether the teaching of English grammar is systematic regarding nominal countability, including articles, quantifiers, subject-verb agreement, etc.

2.2 Chinese learners' difficulties

Despite the tradition of emphasizing the learning of English vocabulary in Chinese education, the need to differentiate the countability of a noun in terms of the meaning used in specific contexts is not adequately taught. The plural meaning in Chinese is implicitly encoded in contexts while the countability of a noun in English will distinctively influence the syntactic environment. This difference between Chinese and English often causes Chinese learners to be unaware of the concept of nominal number in English, resulting in errors with regard to the form of the word in specific contexts. For instance, *experience*, *punishment* and *difficulty* are syntactically countable and uncountable depending on the meaning they take in specific contexts. Furthermore, the memorization, mainly the rote-learning strategy in Chinese education, demands much cognitive effort from Chinese learners, especially regarding irregular plural inflections or nominal variability. Consequently, Chinese learners commit many formal errors (Gui and Yang 2003; Chuang 2005: 26-27).

In short, Chinese learners at college level are aware of nominal countability and are assumed to be familiar with the basic grammatical rules regarding numeral inflections, article use, quantifiers, subject-verb agreement, etc. Their native language (e.g. generally Mandarin Chinese) and culture greatly influence their conceptual understanding of nominal countability. However, as shown in many studies (Chang 1987; Milton 2001; Gui and Yang 2003; Chuang 2005; Tang 2006), their skills in using such consciously-acquired knowledge actively are not at all promising. Despite the early input of nominal countability in English education, Chinese learners constantly show uncertainty regarding the concept of countability, and find it hard to identify whether a noun is countable or whether it takes countable or uncountable senses.

There are various reasons for the great difficulties Chinese learners experience in their productive use of nominal countability. First of all, there are no corresponding function systems in Chinese, including numeral inflections, articles, subject–verb agreement, etc. The great divergence between Chinese and English language and culture demands Chinese learners consciously learning many of the grammatical features obligatory in English but absent from Chinese. Additionally, owing to the heterogeneity in languages and cultures, the conceptual understanding of the countability of the material world in Chinese is greatly different from English. Chinese learners therefore have to adjust their mappings about the lexical, semantic, syntactic and pragmatic features between the target foreign language (English) and their native language (generally Mandarin Chinese). Native speakers may be unaware of judging the countability of a noun, using articles or quantifiers intuitively, but foreign learners like Chinese have to memorize the corresponding grammar rules in this respect and use them consciously to maintain accuracy and fluency in their writing. For instance, *luggage*, *information* and *furniture* are uncountable in English, but countable in Chinese. So, it is common for Chinese learners to write *two/many luggage(s)/information(s)/furniture(s)* or to put an article *a/an* before these words.

Furthermore, there is no clear-cut distinction made between countable and uncountable nouns in grammar books (cf. Svensson 1998: 51-73). Some grammar books have become less definite in defining countability and plurality, blurring the boundaries between the countable and the uncountable according to contextual meanings (Tang 2006: 274). The variability in nominal countability also occurs as a result of social developments. Some traditionally uncountable nouns can be used as countable nouns acquiring new meanings in different disciplines, such as *moneys* in the business sense. The word *mail* is uncountable in traditional contexts, but *a mail* is appropriate when *mail* refers to the concept of an e-mail. On the other hand, grammar books and English teaching fall far short from covering the deviations with which foreign learners are confronted in their actual production activities.

As Svensson (1998: 13) states that "number is not only reflected in the nouns themselves, but also in verbs, modifiers and pro-forms", the misconception of nominal number may result in various errors in production activities. Common errors by Chinese learners related to nominal countability are the following: wrong inflection forms, misuse of the articles *a/an* and *the* or no article before a noun, mixing countable with uncountable nouns, wrong word forms in contextual usage, and subject–verb disagreement especially involving collective nouns. Among these errors, those caused by the misuse of articles rank high in writing (Gui and Yang 2003; Chuang 2005). A frequency count (Sinclair 1991) indicates that in English the definite article *the* is the most common word and the indefinite article *a/an* is the fifth most common, so any difficulty with articles tend to be very conspicuous in foreign learners' production activities. The errors caused by the misuse of articles or quantifiers rarely lead

to miscomprehension. However, they will affect the exactness highly valued in written work, as Master (1997: 216) points out: "[i]mperfect control [of using articles] may consciously or unconsciously, suggest imperfect knowledge, and if such an attitude is engendered in the mind of a professor grading a student's paper, that student's grade may suffer."

2.3 Chinese learners' information needs

When the profile of Chinese learners and their difficulties incurred by nominal countability is taken into consideration, Chinese learners expect to find the relevant information explicitly and explicatively presented in learner's dictionaries for their ready use in production activities. Therefore, it is significant for lexicographers to identify foreign learners' real information needs for lexicographic assistance covering nominal countability. Chinese learners' information needs concerning nominal countability can be rendered in the following specifications:

- (a) the articles or quantifiers before a noun,
- (b) the plural forms of the countable noun,
- (c) the usage of a noun used both as countable and uncountable,
- (d) the specific usage of the singular or plural form of a noun,
- (e) the verb form following a noun, and
- (f) the countability of neologisms.

These information needs are influenced by foreign learners' native language and culture because foreign learners with different language and cultural backgrounds may have general as well as specific difficulties regarding nominal countability. Nonetheless, learner's dictionaries as well-acknowledged reference tools in foreign language learning are expected to address these problems by presenting the necessary data as transparently as possible for the convenience of foreign learners. A closer look at current printed and online learner's dictionaries in the following sections is meant to examine to what extent present lexicographic practices achieve their claimed pedagogical functions of making provision for learners' information needs in foreign language learning.

3. Problems with the present lexicographic practice with nominal countability

In the process of searching for lexicographic assistance for communicative purposes, dictionary users normally experience two stages: access to the needed data and retrieval of the necessary information from the corresponding data.

Thus, two essential questions are relevant to the presentation of the lexicographic data the target users need in their particular situations: is the required data easily identified and quickly accessed; and is the needed data comprehensible for the information retrieval with a minimum effort? Simply put, foreign learners may confront access-related and comprehension-related difficulties in the process of dictionary consultation when they turn to dictionaries for information related to nominal countability. An analysis of the present lexicographic practice will be illuminating.

The latest major English learner's dictionaries in print and in online format indicate nominal countability by labels [C] and [U]. The specifications like [singular], [plural], [often plural] [usually plural] and [U/C+ singular/plural verb] are supplemented in some cases. Abbreviations like [C], [U], [sing.], [pl.] are normally used, for reasons of space, in the latest printed dictionaries, such as *Oxford Advanced Learner's Dictionary of Current English* and *Macmillan English Dictionary* (henceforth OALD7 and MED2 respectively). Online dictionaries, free of space restraints, generally indicate such information unabbreviated. Usage examples are usually provided following the sense differentiation. To sum up: The data relevant to nominal countability are normally presented in the form of labels and specifications before definitions, and demonstrated in usage examples, regardless of the dictionary formats. As discussed above, this lexicographic practice has seldom been challenged. However, is such a lexicographic practice effective to address learner's difficulties related to nominal countability in production-oriented activities? In the following discussion, four articles **education**, **effort**, **fish**, and **information** are randomly taken from a printed dictionary, OALD7, and online dictionaries, such as *Longman Dictionary of Contemporary English Online* and *Macmillan Dictionary Online* (henceforth LDCEO and MDO respectively) for a close examination. In Chinese, the concepts *education* and *effort* are abstract and understood as uncountable, and the concepts *fish* and *information* as countable. In writing, the possible deviations Chinese learners make with regard to these four words could be: *get education*, *receive good education*, *make (an / many) effort(s)*, *a / many fish(es)*, *an / many information(s)*.

3.1 Absence of indicating articles or quantifiers used before a noun

Labels [U] and [C] indicate the countability of a noun, but fail to provide explicit information for foreign learners' use of articles or quantifiers before the noun, especially for Chinese learners whose native language lacks the corresponding grammar function systems. Moreover, whether the article *a* or *an* should be used before a countable noun depends on pronunciation, not spelling. It is even more difficult for Chinese learners to be aware of the use of no article or a definite article *the*. Therefore, they may insert articles when unnecessary, but omit articles when necessary. They have to consider which article is appropriate to be used before a noun even after having consulted dictionaries.

Even the label [U] may not prevent Chinese learners from producing *three/much information(s)* if they are also uncertain about quantifiers. To some extent, the labels [U] and [C] are testing learners' grammar knowledge rather than assisting them.

3.2 Equivocal indication in the countability features of a noun

Marking a noun as both countable and uncountable without further specification would confuse Chinese learners rather than assure them of the correct usage in production activities. For instance, **effort** is marked as countable and uncountable when it means "attempt" (e.g. LDCEO and MDO). Does this mean that *make an/the effort/efforts* and *make effort* are grammatically appropriate? In LDCEO all examples following this sense only show the countable features, ignoring the uncountable features, either intentionally or unintentionally. The treatment in MDO is almost similar except that the phrases *make little/no effort* are provided, but the plural form *efforts* is used in the example demonstrating the usage of the phrases. In contrast, in OALD7 **effort** is marked as countable when it means "attempt". Does this indicate to foreign learners that **effort** can be syntactically singular or plural when they intend to use it in the above-mentioned contexts? If so, the marking of countable and uncountable seems to be superfluous.

Moreover, the bracketed specifications like [often plural] and [usually plural] tend to be ambiguous, as there is always uncertainty about how to interpret the word "often" and "usually" in production situations. The targeted Chinese learners, still in the process of developing their foreign language literacy, are rather "passive" in using their knowledge about English, and expect explicit guidance from the dictionaries rather than information that demands further interpretation. The authoritative status dictionaries have in the Chinese learning culture just strengthens such expectations.

3.3 Discrepancy in indicating information relevant to countability

The specifications [singular] or [plural] help learners clarify their perceptions about countability of a noun in their immediate situations, but the combination of the labels [C] or [U] with these specifications may cause confusion. For instance, in LDCEO as well as in OALD7, **education** is marked as [singular, uncountable] when it means the process of teaching and learning or a particular kind of teaching or training. However, being uncountable already means the word cannot be used grammatically in plural and will only occur in its uninflected form in usage, because "an uncountable noun has only one form, not a separate singular and plural" (OALD7: R42). English learning in Chinese educational contexts strictly observe such an understanding as the concept of countability is a key point examined in English tests at various levels. There-

fore, such marking seems to be superfluous and may be misunderstood by Chinese learners.

Additionally, the discrepancy between the labels and usage examples indicating the countability feature of a noun tends to confuse Chinese learners. For instance, in LDCEO the word **education** has been explicitly marked as uncountable when it occurs in senses 1, 2 and 3, but the phrase examples *get/receive an education* are provided under the first sense. This therefore contradicts the definition on the back of the front cover of MED2 "uncountable nouns that cannot be used with *a* or *an* or a number and have no plural". Being uncountable labels it as impermissible in combination with *a/an*. This matches Chinese learners' experience in English classrooms in China. Faced with such data presentation in dictionaries, Chinese learners may be wondering which rule to follow. Considering the authoritative status of dictionaries in Chinese learning culture, Chinese learners may follow the examples in their immediate situation, but their uncertainty will not be settled.

3.4 Inefficiency of exemplifications

Examples have long been considered an effective way to demonstrate syntactic behaviour in context. However, the potential of examples is not fully exploited for demonstrating the information related to nominal countability, especially regarding those nouns with which foreign learners tend to commit errors. For instance, the example from LDCEO, *I need more information*, is not effective to show that *much information* instead of *many information* is appropriate. Similarly, in LDCEO, MDO and OALD7 **fish** has been marked as countable when it means "an animal living in water which swims", but the listed examples do not assure Chinese learners of the appropriateness of their writing *three fish(es)*.

3.5 Difficulty of accessibility

In both printed (e.g. OALD7) and online dictionaries (e.g. MDO), the accessibility to the needed data still demands much effort. In printed dictionaries, users have to carefully scrutinize the heavily-packed entry to locate the needed data. Also, the online dictionary MDO does not show much improvement in this respect, although theoretically there are no space restraints for online dictionaries. The thin font and the crammed data make it hard for users to quickly access the needed data, adding mental pressure to them as well. LDCEO performs the best among these dictionaries, but still remain traditional in directing users to the required data. If users start their search with *a/an information*, LDCEO will produce an indifferent comment like "a information has returned no result" rather than direct users to the article **information** with the correct expression highlighted for easy identification, while MDO directs users to the phrase *a mine of information* in the article **mine** instead of *a piece of*

information or *much information* in the article **information**. Moreover, some nouns whose meanings in singular and plural forms are different from each other are still condensed in the entries of their singular forms, such as *glasses* (referring to spectacles) in the entry **glass**, and *arms* (referring to weapon) in the entry **arm**. Such presentations tend to prolong the access time.

As can be seen from the above analysis, the present lexicographic practice regarding countability demands much effort from learners in the process of data access and information retrieval. The deficiency in the present lexicographic practice can be attributed to the native dictionary-makers' neglect of the target users' specific needs in a concrete situation (e.g. Chinese learners in English writing), or their overestimation of foreign learners' knowledge of English vocabulary and grammar. In real situations, foreign learners consult dictionaries to retrieve the needed information for their immediate use and have no interest in delimiting linguistic concepts of nominal countability. Dictionaries are expected to actively accommodate their lexicographic needs arising in concrete situations rather than being a repository of knowledge waiting to be exploited. Béjoint (1981: 221) proposes that "foreign students need to be given enough information to help them avoid mistakes, and possibly even to attain ease, elegance and subtlety in their use of the foreign language". But lexicographers should go further by explicitly presenting the necessary information for quick access and easy retrieval by these users in their particular situations.

4. Suggestions for presenting the required data regarding nominal countability

In view of the deficiency in the present lexicographic practice of presenting nominal countability, and the profile of Chinese learners, the following proposals are made for future lexicographic research and practice in this respect. These suggestions are free from dictionary formats, and adaptations may be needed with regard to the practical aspects when it concerns the different dictionary formats and different dictionary user groups. The example articles are presented after each suggestion for demonstration purposes. In order to highlight the data concerning this study, other lexicographic data in the example articles are omitted on purpose. Some usage examples are taken from dictionaries examined in this study.

4.1 Presenting information on articles and the plural forms of nouns

To assist foreign learners in immediate usage situations, the articles *a/an* and *the*, or a dash to indicate no articles, as well as the plural inflection of a noun, such as *-s* or *-es*, can be bracketed before the definition part when the noun is marked as countable. If the plural inflection is irregular, the full word form should be presented. (This has been practised in some learner's dictionaries.) It

is also better to indicate that the uncountable nouns cannot be used with the indefinite articles *a/an* or cardinal numbers and that they undergo no formal changes in terms of number. In print dictionaries, the swung dash (~) can be used to indicate the headword for reasons of space, while in online dictionaries, it is recommendable to display the full word form to avoid any possible confusion. Such treatment will save foreign learners much effort to consider whether an article is needed or which article should be used as well as the possible plural inflections. On the other hand, it will raise the awareness of foreign (e.g. Chinese) learners regarding the use, when necessary, of articles, cardinal numbers or plural forms. Given the fact that foreign learners may find it difficult to process linguistic metalanguage, this study argues that it is necessary to clearly indicate the articles used before a noun rather than to use grammatical terminologies like "no indefinite articles".

Being limited, this study does not address learners' difficulties with quantifiers used before nouns, but it holds that more explorations should be conducted in this respect. In some of the following articles, however, a few examples with quantifiers will be given, but the use of quantifiers before nouns needs careful consideration. For instance, some nouns, like *bread*, *noodle*, and *vegetable*, cannot be combined with quantifiers with one contextual meaning, but when they assume another contextual meaning, it is possible. When these nouns assume the contextual meaning of variety, they could be combined with articles *a/an* or quantifiers *two* and *many*. When they refer to the food as a whole such combination is impossible. Example article 3 *noodle* (below) demonstrates this.

oasis /.../ **noun** (an oasis, the oasis; plural: oases, the oases) ...

Example article 1

bread /.../ **noun**

1. (Use: **bread**, not: ~~a bread~~, ~~two breads~~) a type of food made from ...
a piece/loaf of bread
2. ...

Example article 2

4.2 Delineating the usage of nouns that are both countable and uncountable

If a noun is countable, having both singular and plural forms, but tends to be used in the plural form, such as *noodle*, *congratulation* (when it refers to the words used to congratulate someone), the plural form should be provided either by presenting the inflectional suffix, such as *-s* or *-es*, or showing the full word form if there is irregular numeral inflections. (This has been widely prac-

tised.) Explicit recommendations should be given concerning the circumstances under which the singular or the plural form of a noun is used, instead of vague expressions like "often" and "usually", as it is difficult for foreign learners to decide whether their usage situations can be categorized into the "often" or "usually" cases. The recommendations can be indicated in different ways: stated directly; transmitted through the definition; demonstrated in phrase or sentence examples.

noodle /.../ **noun** (plural: **noodles**) ...

When we refer to food, we use *noodles*.

When we refer to the physical object, we can say *a noodle, many noodles*.

a bowl of noodles

Chinese food is often served with rice or noodles.

Example article 3

congratulation /.../ **noun** ...

1. When you congratulate someone, you should say **congratulations**, not **congratulation**

congratulations on

2. (Use: **congratulation**; not: **congratulations**) the act of congratulating sb
a letter of congratulation

Example article 4

A distinction should be made with explanations or typical examples when a noun is used both as countable and uncountable. There are two possible situations when a noun is marked as countable as well as uncountable. If the countability features of a noun are connected with its meanings in context, such as *experience, difficulty, fruit, vegetable, room, etc.*, the distinction should be illustrated explicitly in definitions, by synonyms or even equivalents (e.g. Chinese equivalents for Chinese learners), when necessary. Moreover, the examples immediately following are expected to help foreign learners decide on the appropriate word form in the relevant syntactic environments. On the other hand, if a noun can be used both as countable and uncountable with the same contextual sense, recommendations or prescriptions are needed, depending on the dominant countability features in the specific contexts. For instance, the noun *effort* is prominently used as a countable taking both singular and plural forms in contexts when it means "an attempt to do something", despite being marked as both countable and uncountable. Therefore, there is no need to add to foreign learners' uncertainty by indicating this word as having dual identities of countability.

experience /... / **noun** ...

1. (Use: **experience**) 经验

*She has gained much **experience** in the computer industry.*

2. (Use: **experience**) (生活) 阅历

***Experience** has taught me that life can be very unfair.*

When **experience** refers to 经验, (生活)阅历, you can only use:
experience + singular verb, not: ~~**an experience**~~, ~~**many experiences**~~

3. (a/an + (adj) + **experience**; plural: **experiences**) 经历, 体验

*an unforgettable/exciting/enjoyable **experience***

*our childhood **experiences***

*Living in Africa was very different from home and quite **an experience**.*

Example article 5

effort /... / **noun** ...

1. ...

2. (**an effort**; **efforts**) an attempt to do something

*an **effort** to do something*

*make an/every **effort** to do something*

*make **efforts** to do something*

Example article 6

4.3 Demonstrating information about subject–verb agreement

The difficulty of foreign learners with the subject–verb agreement in connection with nominal countability should be treated carefully, especially when collective nouns are involved. Generally speaking, the problems caused by the understanding of collective nouns can be classified into the following cases: (i) the collective noun takes the singular word form and can be syntactically singular and plural depending on the intended meaning when used in a particular context, such as *class, family, team, audience*, etc.; (ii) the collective noun takes the singular form but can only be used with a plural meaning and cannot be used together with the indefinite articles *a/an* or the singular quantifiers *each* or *every*, such as *cattle, people, police*, etc.; (iii) the collective noun takes a plural form and is syntactically plural, such as *clothes, glasses, goods, belongings, credentials*, etc., but cannot be preceded by cardinal numbers (e.g. ~~*two clothes*~~) or have the singular form maintaining the meanings unchanged; (iv) the collective noun has the singular form and is syntactically singular, such as *luggage, furniture, equipment*, etc., but cannot be used with the indefinite articles *a/an* or cardinal numbers. However, even the widely-acknowledged dictionary MED2 seems to ignore foreign learners' difficulties with the subject–verb agreement

caused by their misconception of or uncertainty about the collective nouns. For instance, it is almost impossible for foreign learners to abstract information about the appropriate verb forms following the word **cattle** either from the label [pl] or the examples *diary cattle* and *a cattle ranch* in the article **cattle**. This study therefore proposes the corresponding treatment of collective nouns as follows: indicating possible articles when necessary, giving explicit usage notes on verb forms, presenting demonstrative examples and providing common error warnings.

cattle /.../ **noun** (Use: **cattle** + plural verb, the cattle, not: ~~a cattle, cattles~~)

...

The **cattle** are grazing.

Hundreds of **cattle** are slaughtered every day.

Example article 7

family /.../ **noun** (a family; the family, plural: **families**) ...

Use: **family** + **singular verb** when referring to a whole family

My family has always been close.

Use: **family** + **plural verb** when referring to family members (BrE)

The whole family have colds.

Example article 8

In addition to collective nouns, there is a category of nouns ending in *-s* or *-es*, such as the names for certain diseases (e.g. *measles, mumps, arthritis*), sciences (e.g. *physics, acoustics, economics*), games (e.g. *marbles, skittles, draughts*), and countries (e.g. *United States, Netherlands*). Despite the seemingly plural endings, these words are used as singular in syntactic environments. The presentation of the data concerning the use of the articles before these words and the subject-verb agreement can follow the above suggestions in accordance with the context in which the word is used. It should take into account that the contextual meaning of these words may influence their syntactic behaviour, for instance, *marbles* can be counted (e.g. *many marbles*) when referring to the round objects of the game instead of the game itself. Thus, the presentation of the relevant data is expected to be adapted accordingly.

4.4 Using examples and error warnings to raise learners' awareness

Examples should be amply employed to demonstrate the countability features of nouns, especially those words with which foreign learners tend to commit errors, for instance, using articles before words with deceptively plural endings (e.g. *news, means*) and the following verb forms in relevant syntactic environ-

ments. Others like *advice* and *information* are uncountable and invariable in their base forms, but foreign learners with certain language backgrounds tend to treat them as countables. Examples like *The news is good* and *The police have arrived* will be indicative of the syntactic behaviour of these nouns and help learners retrieve the needed information more easily. When necessary, warnings on common errors made by foreign learners can be provided for drawing learners' attention to the appropriate word forms. The error warnings can take different forms depending on the nature of the errors and the profile of the foreign learners. Both the examples indicating the salient countability features of a noun and the highlighted errors should be understood from the perspective of the foreign learners.

advice /.../ noun ...

My supervisor gives me some good advice in my research.

His advice on my study is always enlightening.

Error warnings

You can say/write: a piece of advice, two pieces of advice, a lot of advice,
not: ~~an one advice, two advices, many advices~~

Example article 9

4.5 Integrating study pages on English grammar

Integrating study pages on specific grammatical points in learner's dictionaries, such as nominal countability may be helpful for foreign learners to study certain aspects of English grammar systematically in their cognitive learning situation, but not quite relevant to foreign text production activities. In production-oriented situations, fewer learners will consult this information as they aim to abstract specific grammatical information about individual words in their immediate situations (e.g. article use) rather than the grammar in general. Therefore, this study restricts its discussion in this respect. Admittedly, the study pages are necessary if the dictionary is aimed to serve multifunctional purposes, but the point of departure should be its intended users' grammar knowledge and their needs in the particular user situations rather than the linguistic knowledge itself (Andersen 2007: 133).

4.6 Making an improvement in data accessibility and information retrieval

In view of the above analysis, the presentation of the data related to nominal countability is expected to be explicit and salient for the purpose of quick data accessibility and easy information retrieval. Typography can be fully used for easy data identification. For instance, the bracketed data shaded in the above

example articles can be highlighted in colour in real reference works. The colour font is abandoned in the example articles because it cannot be rendered here. Online dictionaries should exploit the full potential offered by the development of technology, especially regarding search and access routes. For instance, the search starting with a misconception of nominal countability, such as *an advice*, can direct foreign learners to locate the appropriate expressions in the related articles. Compared with the online dictionaries, the printed dictionaries are restrained in space and search routes. But as shown above, it is possible for printed dictionaries to present the necessary data clearly giving entry status to those nouns whose meaning in their plural forms differs greatly from that of their singular forms and place them adjacent to each other instead of strictly observing the alphabetical order, such as *glass/glasses, good/goods, arm/arms, custom/customs*, etc. According to the profile of the learners, the access routes to the needed data can also be various.

The usage notes and error warnings in the box are mainly written in English for the benefit of the readers of this article. In real situations, foreign learners' native language may be considered depending on the quality of the concerned data and the profile of the dictionary users. This study is intended to optimize lexicographic solutions to foreign learners' problems regarding nominal countability, but the realization of these solutions depends on the real situations. This study takes into account that more comprehensive future research is needed, for instance the presentation of information concerning quantifiers. Neologisms should also receive their due attention, as they are concurrent with social development. Many new words have been coined to describe the dynamic world of the information age with some words passing beyond their traditional sense, for instance *blog* is replacing *diary*, and *mail* is referring to *email*. Thus, the countability features of some nouns need to be reconsidered in lexicographic practice.

5. Conclusion

This study shows that the current lexicographic practice regarding nominal countability does not effectively help Chinese learners out of their difficulties with nominal countability in English writing. Foreign learners' lack of reference skills or unfamiliarity with dictionaries cannot in this respect justify the deficiency in dictionaries. Admittedly, the dictionaries discussed target ESL/EFL learners generally rather than Chinese learners specifically, which could partially account for the deficiency in the lexicographic treatment of nominal countability as shown in this study. However, this fact just accentuates the need to create learner's dictionaries that are actively involved in foreign language learning by giving serious consideration to the specific group of foreign learners and the particular situations that prompt their dictionary consultations. It is essential that the presentation of the lexicographic data be understood from the target dictionary users' perspective. Different lexicographic data

should be integrated and balanced to achieve the pedagogical function intended by learner's dictionaries.

It is unrealistic to expect foreign learners in the immediate situation to analyze and synthesize linguistic data covered in dictionaries as it demands them to have full knowledge about the vocabulary and grammar in the target language. It is the lexicographers' role to identify the real lexicographic needs of the target users in lexicographically-related situations, utilize linguistic expertise to make analytic and synthetic work on the corresponding data, and present the necessary data for quick data access and easy information retrieval from the target users' perspective. Upon minimizing dictionary users' frustrations during their look-ups, it is still meaningful to consider Rundell's (1999: 48) proposal a decade ago: "[a] more realistic strategy is to aim for dictionaries whose structure is so transparent that students do not need to *learn* how to use them, and whose content is presented in such straightforward terms that users will have no difficulty in grasping it."

References

- Andersen, B.** 2007. Dictionary Grammars. *Hermes, Journal of Language and Communication Studies* 38: 119-136.
- Béjoint, H.** 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2(3): 207-222.
- Chang, J.** 1987. Chinese Speakers. Swan, M. and B. Smith (Eds.). *Learner English*: 224-237. Cambridge: Cambridge University Press.
- Chuang, F.Y.** 2005. Article Misuse: A Neglected Problem in Chinese EAP Student Writing. M. Singhal and J. Lontas (Eds.). 2005. *Proceedings of the Second International Online Conference on Second and Foreign Language Teaching and Research: Initiative, Innovation, and Inspiration*: 25-33. The Reading Matrix Inc., United States.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Gui, S.C. and H.Z. Yang.** 2003. *Chinese Learner English Corpus*. Shanghai: Shanghai Foreign Language Education Press.
- Hausmann, F.J. and A. Gorbahn.** 1989. COBUILD and LDOCE II: A Comparative Review. *International Journal of Lexicography* 2(1): 44-56.
- Hornby, A.S.** 1965. Some Problems of Lexicography. *English Language Teaching* 19(3): 104-110. *Longman Dictionary of Contemporary English Online* (LDCEO). <http://www.ldceonline.com/> *Macmillan Dictionary Online* (MDO). <http://www.macmillandictionary.com/>
- Master, P.** 1997. The English Article System: Acquisition, Function, and Pedagogy. *System* 25: 215-232.
- Master, P.** 2002. Information Structure and English Article Pedagogy. *System* 30: 331-348.
- Miller J.** 2006. An Investigation into the Effect of English Learner's Dictionaries on International Students' Acquisition of the English Article System. *International Education Journal* 7(4): 435-445.
- Milton, J.** 2001. Elements of a Written Interlanguage: A Computational and Corpus-based Study of Institutional Influences on the Acquisition of English by Hong Kong Chinese Students.

- James, G. (Ed.). 2001. *Research Reports, Volume Two*. Hong Kong: Language Centre, Hong Kong University of Science and Technology.
- Palmer, H.E.** 1938. *A Grammar of English Words*. London: Longmans, Green & Co.
- Robertson, D.** 2000. Variability in the Use of the English Article System by Chinese Learners of English. *Second Language Research* 16(2): 135-172.
- Rundell, M.** 1999. Dictionary Use in Production. *International Journal of Lexicography* 12(1): 35-53.
- Rundell, M. (Ed.)**. 2005². *Macmillan English Dictionary*. New York: Palgrave Macmillan.
- Sinclair, J.** 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Svensson, P.** 1998. *Number and Countability in English Nouns*. Uppsala: Swedish Science Press.
- Swan, M. and B. Smith.** 1987. *Learner English*. Cambridge: Cambridge University Press.
- Tang, E.Y.** 2006. *Linguistic Environment as a Determinant of English Countability*. <http://www.paaljapan.org/resources/proceedings/PAAL11/pdfs/22.pdf>: 266-280. [November 4, 2009.]
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Wehmeier, S. (Ed.)**. 2005. *Oxford Advanced Learner's Dictionary of Current English*. Seventh Edition. London: Oxford University Press.

Notes on Compiling a Corpus-Based Dictionary*

František Čermák, *Institute of the Czech National Corpus, Faculty of Philosophy, Charles University, Prague, Czech Republic*
(frantisek.cermak@ff.cuni.cz)

Abstract: On the basis of sample analysis of a Czech adjective, a definition based on the data drawn from the Czech National Corpus (cf. Čermák and Schmiedtová 2003) is gradually compiled and finally offered, pointing at the drawbacks of definitions found in traditional dictionaries. Steps undertaken here are then generalized and used, in an ordered sequence (similar to a work-flow ordering), as topics, briefly discussed in the second part to which lexicographers of monolingual dictionaries should pay attention. These are supplemented by additional remarks and caveats useful in the compilation of a dictionary. Thus, a brief survey of some of the major steps of dictionary compilation is presented here, supplemented by the original Czech data, analyzed in their raw, though semiotically classified form.

Keywords: MONOLINGUAL DICTIONARIES, CORPUS LEXICOGRAPHY, SYNTAGMATICS AND PARADIGMATICS IN DICTIONARIES, DICTIONARY ENTRY, TYPES OF LEMMA, PRAGMATICS, TREATMENT OF MEANING, POLYSEMY, CZECH

Opsomming: Aantekeninge oor die samestelling van 'n korpusgebaseerde woordeboek. Op grond van 'n steekproefontleding van 'n Tsjeggiese adjektief, word 'n definisie gebaseer op data ontleen aan die Tsjeggiese Nasionale Korpus (cf. Čermák en Schmiedtová 2003) geleidelik saamgestel en uiteindelik aangebied wat wys op die gebreke van definisies aange-tref in tradisionele woordeboeke. Stappe wat hier onderneem word, word dan veralgemeen en gebruik in 'n geordende reeks (soortgelyk aan 'n werkvloeiordering), as onderwerpe, kortliks bespreek in die tweede deel, waaraan leksikograwe van eentalige woordeboeke aandag behoort te gee. Hulle word aangevul deur bykomende opmerkings en waarskuwings wat nuttig is vir die samestelling van 'n woordeboek. Op dié manier word 'n kort oorsig van sommige van die hoofstappe van woordeboeksamestelling hier aangebied, aangevul deur die oorspronklike Tsjeggiese data, ontleed in hul onbewerkte, alhoewel semioties geklassifiseerde vorm.

Sleutelwoorde: EENTALIGE WOORDEBOEKE, KORPUSLESIKOGRAFIE, SINTAGMATIEK EN PARADIGMATIEK IN WOORDEBOEKE, WOORDEBOEKINSKRYWING, SOORTE LEMMAS, PRAGMATIEK, BEHANDELING VAN BETEKENIS, POLISEMIE, TSJEGGIES

* This article is an edited version of a plenary address delivered at the conference on 'Dictionaries, More than Words', which took place at the Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia, 6 February 2009.

Introductory Remarks

The following notes discuss some of the problems and issues encountered during the compilation of a monolingual dictionary. After some preliminary remarks, these notes are split into two main parts. In Part I (2–4), an example and discussion of an analysis of corpus data (drawn from Czech) is offered, resulting in a lexical profile of a word. In Part II (5–8), building on the previous part, a commented list of some of the main aspects and principles of the dictionary-making process is presented. References to most of the points raised are to be found in the literature at the end.

Before any work can begin, a series of decisions laying down the ground rules must be made. Some of these are quite straightforward, others may be more difficult. The major ones may briefly be listed as follows:

Firstly, (a) resources have to be decided on. These include a large balanced (representative) corpus, in addition to some secondary resources, if available. Here it is necessary to be rather wary of the Internet, which is not always as generally useful as is often presumed.

Next, (b) the type of dictionary has to be decided. In the present context, such a dictionary should have, among others, the attributes monolingual, large, synchronic, representative and descriptive (not prescriptive).

Connected with this, though not necessarily dependent on it, is to make a decision about (c) the target users, for example the general public as well as a specialized public.

Finally, taking account of the shortcomings of existing dictionaries and the advantages of having a corpus, a major decision should be made, to be applied in various forms, about (d) the main concerns and orientation of the proposed new dictionary. Three general aspects should be stressed: syntagmatics, usage and context. This implies that attention will be paid to all relevant variants of words and phrases in a language, a fact that is often underestimated or even disliked by prescriptive normalisers and codifiers of a language.

Some attention should be given to (e) software. Here, one of the obvious candidates available today is represented by TshwaneLex. The choice should be made with a view to possible re-use of the data, expansion, redesign, additional products, etc.

PART I

1. Data and Treatment

Any dictionary aims — or should aim — at a true mapping of the meaning behind words, though modern dictionaries include mapping of their use, too, which is made possible by modern corpus data. A large dictionary should respect these data as best it can: it will basically be a corpus-driven product. A major new problem for lexicographers is the profusion of data that was not

available until only a few years ago. For this reason, ways and means are still being sought how to best handle such abundant data. Special and constant problems are identification of the meaning behind corpus forms, splitting the corpus data into manageable groups, interrelationships and elegant and efficient description. By way of analysis of a given lemma, it will, hopefully, be possible to point to some central problems of description and explanation of the meaning. Differences between the traditional approach and the corpus approach can best be illustrated if both the description of a given lemma, taken from an existing dictionary, and a description compiled on the basis of corpus analysis are compared in some detail. As a rule in theoretical models of data analysis, attention is paid almost exclusively to verbs and their salient formal and semantic features, following the preoccupations of syntacticians, while almost no attention is paid to nouns, the largest of the word classes. The latter are seriously in need of more detailed inspection. In addition, some attention must also be paid to adjectives, which stand between the two poles.

2. Traditional Approach and a Critique

First a case may be considered where a pair of related traditional dictionaries (SSČ and SSJČ) and corpus evidence differ widely. The example selected is the Czech polysemous adjective *měkký* (which corresponds roughly to English *soft*). The portrait of this adjective in dictionaries is rather brief and looks simple and straightforward:

- (1) *poddávající se, málo odolávající tlaku* (yielding or giving way to pressure)
- (2) *vzbuzující (na pohled n. na poslech) dojem jemnosti* (evoking an impression of fineness or tenderness (to the eye or ear))
- (3) *podléhající snadno citu, citlivý, soucitný, povolný* (succumbing to feelings, sensitive, compassionate, compliant)

An inspection of the corpus data (3 549 occurrences in SYN2000 of some 100 million words) — or a sample of it that is presumed to be sufficient — yields a rather different picture. The very first impression one gets is that something is seriously wrong with the dictionary definitions. They do not exhaust the data; they use a problematic metalanguage (employing, among other things, synonyms); and they dissect the semantic continuum in a way that is odd, if not wrong. Specifically, they omit a number of analytical criteria that suggest themselves on inspection of the corpus data. Some of these may briefly be outlined. The first few examples, backed by the corpus, deal with the three meanings given in the dictionaries.

- (a) *Pressure (tlak in definition 1)*. Questions one must ask here include the following:
Is it physical or psychological (abstract) pressure that is meant?

Does *měkká voda* (soft water) yield or give way to pressure? (Hardly!)
What about *měkká norma* (soft norm)? (It does not fit.)
Does definition 1 imply that a soft object may be crumpled, cut, sawed or torn apart? (Not under normal circumstances.)

- (b) Definition 2 relies heavily on the near synonym *jemnost* (fineness, tenderness). This is problematic. The adjective *jemný* is given five meanings in the dictionaries (1 having a smooth surface; 2 graceful/delicate; 3 having a small degree of a quality perceived by the senses; 4 distinguishing exact details; 5 having a specifically high quality). These five meanings are not sufficient to cover the meanings of collocations found in the corpus such as *měkké pohyby* (soft movements), *měkký hlas* (soft voice), *měkká stupnice* (minor key, in music), and *měkké i* (soft i, in orthography). In none of these collocations is the synonym *jemný* correct. Thus, the reference to *jemný* is misleading or useless.
- (c) *Podléhající citu* etc. (feelings, definition 3). Here too, it is difficult to fit this definition to existing collocations, for example *měkký člověk* (soft-hearted man), *měkká povaha* (conciliatory nature), etc. They are different and difficult to describe in this way.
- (d) Next, there is a multitude of examples, illustrated by corpus collocations, that do not fit the definitions either, e.g. *měkký horský vzduch* (soft air), *měkká ekonomika* ('soft economy'), *měkká koncepce* ('soft conception'), *měkké dřevo* (soft wood), *měkké lyže* ('soft skis'), *měkká pornografie* (soft pornography), *měkká radiace* (soft radiation), etc.

3. Some Principles of Corpus Data Analysis

To get to the bottom of the maze of facts that lie behind this adjective and not to leave out anything relevant, a comprehensive and representative (if not exhaustive) concordance must be compiled and analysed. The analysis must be based on random samples, whose number and size will depend on the type and complexity of the lemma being analysed. The analysis that seems to be relevant in most cases consists of a number of steps, mostly manual and rarely simple, always starting from features found in the data (steps 3–5). However, it is necessary to start (steps 3.1 and 3.2) by singling out and setting aside cases that would otherwise complicate the analysis.

3.1 Idioms and Phrasemes

Without going into detail, all of these may be identified on the basis of a paradigmatic or syntagmatic anomaly, which is either semantic or formal in nature. Here, only a few cases are eligible. These include the expressions *mít měkké srdce* (be soft-hearted) and *být měkký na někoho* (to be soft — i.e. not strict — on sb).

Additionally, it should be noted that, although no examples are found in the case of *měkký*, fixed expressions and stereotypical phrases, including catch phrases and proverbs, fall under this heading too.

3.2 Multi-Word Terms

Leaving aside instances of specific terminological meanings of single-word lemmas, which are typical of nouns, there is, in the case of this adjective, some terminology that consists of multi-word terms, such as *měkká voda* (soft water), *měkká droga* (soft drug), *měkká radiace* (soft radiation), *měkká pornographie* (soft pornography), and *měkký konec řádky* (soft end-of-line return).

After this, the gist of the analysis is concentrated in three steps (3.3–3.5).

3.3 Determination of Function

Of the three main adjectival functions, namely (a) attributive only, (b) predicative only and (c) both (majority of adjectives), it turns out that all uses of *měkký* fit into the last type only. Hence no specific functional description is necessary here, though other adjectives may have more specific functionality. Obviously, each word class has one or more specific functions, distinct from the other.

3.4 Semiotic Classification

This largely depends on the part of speech. It is basically pragmatic and corresponds to particular needs. In the present case, it seems sufficient to classify all the nouns qualified by the adjective *měkký* into five broad classes according to the type of denotation of the noun that they modify, namely:

- (a) man (humanus, H), *obchodník (byl) měkký*,
- (b) animal (animalis, An), *krávy jsou měkké*,
- (c) (concrete) thing (res concreta, K), *řízek (byl) měkký*,
- (d) (abstract) thing or abstract (res abstracta, A), *měkká atmosféra*, and
- (e) (place (locus, L), –.

In some cases, it may be useful to identify a sixth class, namely:

- (f) metaphorical use (M), *měkká politika* (literally, soft politics).

This is discussed further under point 3.8 below. Most uses of the adjective *měkký* in the corpus data under inspection fall into (c) and (d).

Only when this analysis is complete, is it viable to look, within these broad classes, for any further markers and features, which may be very important but do not seem to be so general.

3.5 Formal markers

These include any relevant information that the form signals. A desideratum here, though difficult to attain, is to do automatically as much identification as possible of at least the following formal features:

- (a) valency, most prominent with verbs though not limited to them,
- (b) special position or formal use, and
- (c) specific frequent collocates.

While a single valency to be found here is restricted to the idiom mentioned above, no postpositive uses of the adjective *měkký* were found (though some adjectives are so used). Neither was any specific uses of *měkký* with negatives or other special constructions encountered. However, there are some frequent and obvious cases of *měkký* found collocating with *být* (to be), which should be duly noted.

At least two more systematic criteria should be applied in any analysis of corpus data for a lexical item. These are the paradigmatic set membership of the item and its frequency.

3.6 Set Membership in a Collocational Paradigm

By this, the whole range of regular collocations of the item is meant, with the exception of idioms and multi-word terms, though these are closely related. It is no paradox to view a set of collocations, i.e. syntagmatic feature, as a collocational paradigm (one or more). This has not yet been done systematically in any dictionary. However, it gives vital information about the possible uses of the item in text, so it is crucial to mention this kind of information. For practical purposes, this becomes of greatest importance in those cases where the collocational set (paradigm) is comparatively small, restricted to only a few members (i.e. a closed paradigm set). Although no such restricted collocational sets are to be found in the case of *měkký*, the point can easily be illustrated by a different word, the Czech adverb *dokořán*, which is translated as 'fully' in bilingual dictionaries. The fact is, however, this word collocates with only six other words (*otevřít, být, nechat, zůstat; okno, dveře*, i.e. open, be, leave, remain; door, window). It is, then, far more important to give these six collocations in the dictionary, not trying to determine the meaning of *dokořán* at any cost, for this is not easy to specify (in some cases it corresponds to English 'ajar'). In an attempt to find the meaning in this case, generalizing over a mere six occurrences is linguistically problematic: there may not be a sufficient analogy here. A sufficient analogy is a prerequisite for any judgements about the meaning of a lexical item and its type.

3.7 Frequency

It is almost impossible to overestimate the importance of this feature, which is now so well documented in the corpus, but which users, until now, have had no access to. It helps in many ways, not least by indicating which meaning should be recorded as the first in the dictionary.

Before continuing, two more remarks of a general nature must be made.

3.8 Paradigmatic–Syntagmatic

Though a good corpus may offer many different types of information, handling this information may be somewhat idiosyncratic, depending on the type of dictionary. It is evident that new, corpus-based dictionaries should aim to redress the age-old imbalance in information offered previously. As a generalization, it may be said that these dictionaries, because of the limited supply of data and their main purpose, have largely been skewed towards the paradigmatic aspect, emphasizing classifications of various sorts and determining memberships in classes set up by lexicographers.

With modern corpora, however, it is possible, for the first time ever, to offer syntagmatic information in dictionaries as well, indicating vital information about the usage of words in real texts. In lexicography, this amounts to two things primarily, valency and collocations. Though formal valency (such as the case forms required by prepositions) may not be difficult to pin down and should and can be determined for all word classes (not only verbs, such as *depend on*, *abstain from*), collocations still present a problem. It is not so much a matter of their exact theoretical determination — though linguists take widely different positions on this — but rather a matter of practical selection from the vast quantities of corpus data.

One of the problems created by the profusion of data in modern corpora is that one is pushed, by means of various statistical association measures (such as log-likelihood or MI score) towards what is *typical* only, being offered little or no information about marginal, infrequent and, perhaps, untypical uses, which a large dictionary should record or illustrate too. To view marginal collocations as a limitless string of exceptional, figurative or metaphorical uses is hardly a solution. Instead, potentiality of use should be considered here and instances of isolated marginal use should be double-checked against other resources. No doubt, in some cases, such collocations will turn out to be no isolated or figurative uses, but newly emerging types of standard meaning.

3.9 Pragmatic Uses

Finally, pragmatic uses should be identified and a specific semiotic approach devised. What effect does a particular expression have on the reader or listener,

and under what circumstances? A major feature here is evaluative use, which, as it happens, is often of a negative nature.

4. Lexical Profile of the Adjective *měkký* (soft)

The analysis based on the points raised and briefly explained above, has produced a different profile of the lexeme *měkký* from the one with which was started (Čermák 2007). This profile is shown in what follows (though it could, based on different emphases, take other shapes, too). Even the best dictionaries differ widely from any corpus-based profile. So far, few dictionaries have been based on corpus data, and none in Czech. Obviously, a profile such as the one below, if applied in a printed dictionary, would have to be collapsed into the dictionary's description format. It would, however, be expected to preserve all the distinctions found in the corpus and mentioned here (above and below) and to be made clear for the user. The latter point imposes the constraint of a limited metalanguage vocabulary. It is evident that the syntagmatic aspect is made prominent here, especially in the subsenses (a), (b), (c), etc. A sample of *měkký* (soft) that has been analysed is given in the Appendix. The lexical profile, originally compiled in Czech (see Appendix 1), is given here in English for the benefit of a wider readership.

1. **ABILITY and EFFECT (of a concrete object) that is physical for the agent (animate):** *under the influence of pressure or force, easy to shape, cut, saw or fold; elastic and quite resilient*
 - (a) matter, material, product: *having a smooth surface, pleasant to touch*
 - (b) physical object, product: *rounded, not angular*
 - (c) fruit: *very ripe*
 - (d) meal: *prepared, cooked and ready for eating*
2. **EFFECT (of a concrete or abstract object) that is physical, especially acoustic, visual or tactile, for the receiver (inanimate or animate):** *having a pleasant quality including a fine effect or contrast rather than being sharp or pronounced*
 - (a) voice, sound: *quiet and delicate*
 - (b) rain etc.: *not strong, neither severe*
 - (c) contact, fall, blow: *not violent or intensive*
 - (d) consonant: *pronounced as fricative*
3. **EFFECT (of an abstract object) that is psychological for the receiver (animate):** *being sympathetic, benevolent or even compassionate, and sometimes slightly exaggerated*
 - (a) words, language: *not stern neither angry, conciliatory*
 - (b) a human being in their conduct or expression: *conciliatory in politics or irresolute*
 - (c) norm, judicial decision: *not severe, not principled or consistent*

4. **EFFECT of a concrete or abstract object that is different (from that in 1–3) on the receiver (animate):**
- (a) alcoholic drink or other intoxicating substance: *having a weak effect*
 - (b) market, currency, goods: *losing value*
 - (c) water: *without minerals (and unsuitable, among others, for shaving)*
 - (d) drug: *not addictive*
 - (e) radiation: *weakly penetrating*
 - (f) pornography: *suggestive, rather than explicitly erotic*

Further criteria could be introduced to make the overall picture more detailed, such as distinguishing cases where the concrete and abstract are collapsed. This all depends on the degree of granularity that the lexicographer wants to achieve. Naturally, the more detailed the description gets, the less transparent and organized for the user it becomes. The fourth major class, which is complementary to the first three, covers residual types of meaning and usage, and is often terminological and metaphorical; it could easily be expanded into separate categories.

PART II

Notes on Some Stages and Types of the Lexicographer's Work

5. General and Theoretical Issues

Drawing to some extent on the preceding part, which was more practical in nature, some generalizations will be mentioned in this part. The following notes, more theoretical and often very short, do not aspire to be a systematic and full survey of the problems that lexicographers deal with (see, for example, Hartmann and James (1998), Atkins and Rundell (2008) and Hanks (2009, Forthcoming)).

5.1 The Basic Resource: The Corpus

A good and balanced corpus is today essential for the compilation of a dictionary, but it is sometimes necessary to consult additional resources (such as those mentioned in 5.2), either because more information is required or because corroboration of corpus data is needed.

5.1.1 Word List and Frequencies

- A frequency list of words and lemmas is very useful for many purposes, e.g. for determining the likely complexity of an entry.
- Frequency information should be given for all lemmas.

- The list should include all variants found in the corpus, ordered by frequency.
- All members of a closed class should be included (e.g. names of colours).

5.1.2 Selection for Analysis

- When selecting a sample from the corpus for analysis, it is important to avoid one text genre only, wherever possible, and at all events to avoid relying on a single source, which would be too skewed and likely to result in distortion.

5.1.3 Concordances

- The choice of random samples is necessary, if the data for a particular lexical item is too big.
- A manageable selection in a concordance has its limitations, though ordering it may help to overcome some of these, for example to find formal valency markers, collocations, etc.
- Filters can help in making a further selection, if these are available in the corpus browser.
- Statistical measures may offer additional help in decision making, especially with respect to collocations.

5.1.4 Additional Corpus Tools

- Other tools are available, such as *Word Sketches*, though they do not help in decisions about peripheral phenomena.

5.2 Additional Resources

Should the corpus data not be sufficient, then targeted excerpts or even inquiry through distributed questionnaires in special cases might be necessary. (It very much depends on the corpus composition.)

The Internet is not to be trusted as a source of data, in many cases being skewed and full of hiatuses. Its worst performance is probably in the domain of authentic spoken language and dialogue.

5.3 Types of Lemma (Dictionary Macrostructure)

At least four types of lemma/entry should be distinguished, namely:

- Single-word lemmas: most entries; no grouping is preferable.

- Multi-word lemmas: idioms and terms, problems of selection and identification.
- Technical apparatus: abbreviations, cross-references, etc.
- Specialized types of entry may also be envisaged, for example prefixes and suffixes or suppletion forms having a different alphabetical order (English *went, go*).

5.4 The Entry: Some of its Features (Dictionary Microstructure)

In what follows, the single-word lemma will be specifically commented on.

5.4.1 Form

Form includes a number of familiar items whose treatment depends on the dictionary policy. Here, only a brief summary will be given:

- Lemma, variants (a true description of forms that have actually been recorded, not prescriptive, otherwise it could lead to a never-ending selection).
- Grammar (endings, reference to tables, etc.).
- Pronunciation (differential only, some foreign words).

5.4.2 Style, Register

The dictionary should reflect real usage (in contrast to stylistic theories, which are usually far from the world of real language). Labelling should be kept to a minimum and terms must be designated. As register tends to change rather rapidly, any labelling should be reviewed at the end of the compilation.

5.4.3 Additional Features (optional)

With a large dictionary many options open up, which cannot be given much thought and scope in lesser ones. These may include special sections on:

- Frequency (in some simplified form).
- Synonymy (though this should never be a substitute for meaning definition).
- Etymology.
- Special usage notes (mostly pragmatic, perhaps also historical, including notes on differences between variants too).

5.5 Meaning

Rendering a satisfactory description of an item's meaning is the most important goal of any general dictionary. Only a few basic principles will be mentioned:

- Meaning and use are inseparable because meaning can only be deduced from attested use.
- Meaning can be deduced only from real and sufficient contexts of use.
- Each definition should be self-sufficient, not relying on outside information.
- Each definition should be worded sufficiently, so that it does not fit other entries, i.e. it should be unique.
- Definitions should be based on real data and should hold for all significant occurrences of the form.

5.5.1 Types of Meaning

A distinction should be made between the meaning of (a) terms (see 3.2, 6.2) and (b) standard lexemes; both being further distinguished from (c) pragmatics (such as evaluative function).

5.5.2 Definitions

Except for the COBUILD type of definition, most approaches are basically variations of the mainstream. Some of the salient principles are as follows:

- The basic, classical approach is based on the *genus proximum + differentia specifica* dichotomy, i.e. where possible. In today's terms, this boils down to a closest hypernym and a set of specific necessary features.
- Ostensive, deictical definition is useful (if available), using showing and pointing (though indirectly in most cases) to outside objects and phenomena the word is related to. This may include pictures, charts, etc.
- Relational definitions hold for derivatives, but the semantic relations are not always mechanical and additive. This is a frequent source of misinformation as the derivatives hardly ever reflect the base exactly, e.g. between a noun and a related adjective.
- Often, it is useful to give typical referential nouns (for adjectives) or the type of subject, object, etc. (for verbs). This is directly linked with collocations and other syntagmatic information.
- Function (of grammar words, etc.) is not meaning, nor can it be related to other specific lexemes (by way of collocations, etc.).

- Since function is theory-dependent (e.g. conjunctions and particles depend on a theory of syntax and pragmatics), the relevant theory must be stated explicitly in advance, at least by reference to a particular framework.

5.5.3 Polysemy

Polysemy, universal in language in all of its frequent lexemes, traditionally causes difficulties for the lexicographer, there being no consensus as to how to deal with it (but see the suggestions above, put forward on the basis of the analysis of *měkkij*). At least the following general points can be made:

- The meaning and its parts/senses should be related to form wherever possible (i.e. syntactic use, valency, collocability).
- Discrimination between common usage and terminological phraseology is necessary.

5.5.4 Other Semantic Features

These may be viewed as largely (though not invariably) complementary, including:

- Synonyms may superficially seem to be useful, though ideally the users might expect comment about differences between a synonym and the lemma.
- Opposites (not just plain antonyms) are essential if available, helping the users and orienting them in the lexicon system.
- Hypernyms (not necessarily only the immediate ones) are essential and in fact no definition is possible without them.

5.6 Principles of Meaning Definition

A number of specific principles can be mentioned that relate to the description of meaning. Though commonplace, perhaps, these are worth giving here for they should always be kept in mind. Consider at least the following:

- The unknown (and rare) should be explained in terms of the known (and common). There is an advantage in having a specified metalanguage (e.g. the Longman restricted defining vocabulary of 3 000 common and frequent words), though this has not yet been tried for a large dictionary.
- Context and usage is the main arbiter for the meaning of a word often standing in sharp contrast to preconceived ideas.

- There is no standard size of context to be given in examples; it depends on the nature of the lemma.
- Each definition should be equivalent in its form and wording to the relevant part of speech, enabling a possible substitution in text (for auto-semantic/lexical words). Here, a broad substitution test (substituting the definition for the lemma in relevant contexts) may often be helpful.

Nevertheless, the use of paraphrase in the definition should be unambiguous. An alternative is the COBUILD full-sentence type of definition.

- The definition must not be circular (no defining by mutual synonyms is a solution or description).
- Opposites and contrasting words, if there are any, must be mentioned as these are important links to a complementary lexeme.
- All examples should correspond to the definitions given and should not substitute those parts of it that are not mentioned.
- There is no such thing as a specific isolated meaning: The solution is either to find more examples to make it a standard meaning or to declare the combination to be an idiom. The old idea of exception, preserved in grammar perhaps, can be dissolved into either solution indicated above.
- As much as possible must be fitted into the definition, avoiding metaphorical meanings, perhaps by a double-layered approach (i.e. giving a main meaning plus secondary meanings to each sense).
- The possibility must be considered of giving a (simplified) scientific definition of terms versus standard definition (e.g. defining *salt* as 'NaCl, sodium chloride', as well as 'a white crystalline substance used for seasoning or preserving food').
- Collocational restrictions must be observed: If a lemma is found to collocate with a severely restricted class of collocates only, this must either be explicitly stated or the class must be viewed as a set of 'fixed' collocations (idioms) and the lemma must be taken out of the list as not being in use independently.
- (Morphological) forms, occurring in specific collocations usually, often have a specific meaning, not applicable to the whole lemma, hence they may require special treatment in a section of the dictionary article or in an independent lemma.
- Extended, mostly metaphorical cases of use should be carefully selected, if intended for inclusion, especially with regard to showing possibilities of (current or future) expansion of standard meanings that have been recorded, as an indication of the potentiality of the language.

6. Idioms and Terminology (single and multi-word lexemes).

As both idioms and terminology are a matter of a much more complicated and different type of lexicography (see, for example, Čermák 2007), only a couple of general principles may be mentioned here.

6.1 Idioms

These should be given sufficient definitions, including information about use, the classes of users, and the circumstances under which they are used.

- All idioms should remain unrelated to numbered meanings of a single-word lemma and should receive special treatment, including specification of their pragmatic function.
- At least some idioms/phrasemes could be independent entries.
- Many idioms are pragmatic, specifically evaluative and this information should be explicitly given.
- The problem of their alphabetisation should have a simple and systematic solution (e.g., for word classes: first noun, then adjective, then verb, etc.).

6.2 Terms

Constituting the largest part of any natural language (including numerous multi-word expressions), these should always be defined in consultation with experts, who should also assist in the selection of technical terminology.

- In many cases, terms should be given both an encyclopedic and lexicographic definition; the latter may be shorter.
- There are no self-evident criteria for the choice of terms. Some combination of expert advice and corpus frequencies is needed.
- It may be desirable to distinguish between the terminological and common use of lemmas. (See the discussion of *salt* above.)

Finally, it may be useful to mention briefly some practical issues regarding the whole process of dictionary compilation.

7. Technical Aspects of Compilation

There are very many technical aspects of dictionary compilation. Only two of them will be mentioned here.

- A preliminary database could be useful. If available and preannotated, it will save time during the compilation process, although regrettably there is a danger that it might overlook new data if the corpus is growing during the period of compilation.
- Useful tools include ready-made templates (for data split into homogeneous classes) and a style guide (mainly specifying the sequence of steps to be taken and editorial policy decisions).

7.1 Preparation

Once a word-list is available it is advisable to:

- Split entries into homogenous categories, such as parts of speech and their subclasses: This safeguards homogeneity of compilation. But not all words are easily classified in this way.
- Compile an average and a medium-size entry as a pilot exercise: This will provide valuable experience and a basis for modulation of principles. Obtaining a first idea of what an average entry will be like usually serves as a basis for planning (though the conditions specified in any plan are rarely met and fulfilled in all details).

7.2 Further Steps

Before starting in earnest on compiling the dictionary, it is useful to ensure that the data is as homogeneous as possible. This, among others, means:

- Selecting and extracting idioms and other multi-word units for special treatment (see 3.1 and 3.2).
- Identifying pragmatic words and expressions (i.e. those related to society, addressing the basic question 'How does their use affect people?'): Special descriptions accounting for social use (and abuse) and effect are necessary here.
- Creating lexical profiles as a starting point: A useful tool is *Word Sketches*, but only for the core usage of each word.

7.3 Technical Aids

- Statistical association measures such as MI-score and t-score indicate salient combinations and their types. Sometimes even simple bi-/trigrams are helpful, too. On the other hand, no single association measure yields all the collocations that might be of interest.

- Collocations are scalar, ranging from typical to rare and untypical. (A policy is therefore needed to decide how far to go with the inclusion of these.)

7.4 Control Mechanisms

Some control mechanisms are necessary, designed in particular to ensure that

- the same types of entry are handled similarly throughout (for all members of a class), and
- formal mechanisms (such as punctuation and spacing), references, etc. are styled consistently throughout.

8. Open Questions

A number of open questions remain, depending on the specific procedures used. The following at least may be mentioned:

- Maintaining links with an open corpus, specifically when in need of further or new data.
- Drawing the line, i.e. finding a cut-off point between collocations that are quoted and those omitted.
- Including dialect forms and information about them.

Acknowledgement

The author's thanks go to Patrick Hanks for his suggestions in helping to improve the final text.

References

Dictionaries

- SSČ = *Slovník spisovné češtiny pro školu a veřejnost*, ed. J. Filipec, F. Daneš, J. Machač, V. Mejstřík, 2. vyd. 1994 (1. vyd. 1978). Praha: Academia.
- SSJČ = *Slovník spisovného jazyka českého 1960–1971*. Praha: NČSAV / Academia.

Other Literature

- Atkins, B.T. Sue and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

- Čermák, František. 2007. Polysémie a kolokace: případ adjektiva měkký. Čermák, F. and M. Šulc (Eds.). 2007. *Studie z korpusové lingvistiky 2 Kolokace*: 56-93. Praha: NLN.
- Čermák, František and Věra Schmiedtová. 2003. The Czech National Corpus Project and Lexicography. Murata, M., S. Yamada and Y. Tono (Eds.). 2003. *Asialex '03 Tokyo Proceedings: Dictionaries and Language Learning: How Can Dictionaries Help Human and Machine Learning?*: 74-80. The Asian Association for Lexicography.
- Hanks, Patrick. 2009. Compiling a Monolingual Dictionary for Native Speakers. *Lexikos* 20: 580-598.
- Hanks, Patrick. Forthcoming. *Lexical Analysis: Norms and Exploitations*. Cambridge MA: MIT Press.
- Hartmann R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York Routledge.

Appendix 1: Lexical Profile of *měkký* in Czech

1-SCHOPNOST a ÚČINEK konkrétní fyzický pro agenta (živ): *pod vlivem tlaku n. síly snadný tvarovat, krájet, řezat či skládat, pružný, elastický a neodolný*

a-hmota, materiál, výrobek: *s hladkým povrchem a příjemný na dotek*

b-předmět, výrobek: *zaoblený, ne hranatý*

c-ovoce ap.: *velmi zralý*

d-jídlo: *uvařený, připravený k jídlu*

2-ÚČINEK konkrétní/abstraktní fyzický, zvl. akustický, vizuální a hmatový, na receptora (ne/živ): *mající příjemnou vlastnost zahrnující spíše jemný účinek či kontrast než ostrost, vyhraněnost*

a-hlas, zvuk: *tichý a jemný*

b-děšť ap.: *ne silný ani prudký*

c-kontakt, pád, úder: *neprudký*

d-konsonant: *vyslovovaný jako frikativa, třený*

3-ÚČINEK abstraktní psychický na receptora (živ): *sympatický, shovívavý a soucitný, někdy přehnaně*

a-slova, jazyk: *ne příkrý ani rozzlobený, smířlivý*

b-člověk v jednání/projevu: *smířlivý v politice n. nerozhodný, neprůrazný*

c-norma, rozsudek: *nepřísný, nezásadový*

4-ÚČINEK konkrétní/abstraktní jiný na receptora (živ)

a-nápoj a jiná látka: *působící slabou měrou*

b-trh, měna, zboží: *klesající na hodnotě*

c-voda: *bez minerálů (a nevhodná mj. na holení)*

d-droga: *nenávyková*

e-radiace: *málo pronikavá*

f-pornografie: *spíš náznakově, neexplicitně erotický*

Appendix 2: Sample concordance of the lemma *měkký*, organized semiotically (with annotation)

Concretes

- 14: Tenkrát jsem spal taky na *slámě*, jenže byla <měkkčí> . Tahle tlačí a píchá. Chtělo by to posta K
15: jistě víte z teorie i praxe, jsou *dřeva* tvrdá a <měkkká> . TVRDÁ mají hustá vlákna, a proto se K
18: lidský řev, kolo se přehouplo přes *cosi* <měkkého> , a Prokop se probudil. Nahmatal, že K
19: astné, a na to holštýnský řízek právě dost <měkký> , aby lahodil patru, s dozlatova opečený K
20: io, s hmyzím *soustem* přesně tak velkým a <měkkým> , aby zachutnalo jeho ochmýřené, ro K
21: ulisáci zapomněli pod hradby položit *něco* <měkkého> , aby měla na co dopadnout. Výsled K
23: stane. Sádra se nejlépe rozděluje v *nádobě* <měkké> , buď speciální gumové misce, která je K
24: nu se večer ochladilo, lehce přimrzalo a v <měkkém> , chladném *vzduchu* bylo cítit závan j K
27: ne na Žižkov. *Terén* hřiště U Nisy je zatím <měkkčí> , do neděle však pravděpodobně zmrz K
29: barevných kovech jen stručně : *MOSAZ* je <měkká> , dobře se zpracovává, bývá pěkně žlu K
30: átečníky a pro pokročilé. - Coby softcarver <měkká> , dobře ovladatelná, bezproblémová *ly* K
31: my. Nehty a vlasy *Nehty* novorozence jsou <měkké> , dosahují konečků prstů, často je i pře K
33: tvrdě, jak doufala. Slunce svítilo za mraky. <Měkký> , hedvábný *děšt* padal mezi borovice K
34: šest kilogramů, které naše hlava váží, totiž <měkké> , hlavně pak vysoké *podušky* vůbec ne K
35: stavil a člověk mohl pozorovat pohyb jeho <měkkých> , jakoby vycpaných *tlapek*, to jak se j K

Abstracts

- 25: štůje syntezátory a rozeznává jimi zejména <měkkou> , chrámově varhanní *atmosféru*. Svě h A
28: u, pak v za jeho drsnou slupkou objevíme i <měkké> , dobré *jádro*, pak vycítíme, že za jeho A
37: osudu nebylo pouhou náhodou, že by jeho <měkké> , jemné, nehmotné *jméno* odmítalo sp A
41: e ho zmocňuje *cosi* nevýslovně obrovského, <měkkého> , lehkého, průsvitného a přečistého A
47: radace je charakterizována stupnicí : velmi <měkká> , měkká, měkkí, normální, tvrdší, tvrd A
48: . alternativních scénářů, nabízejících jakýsi <měkkčí> , mírnější, ohleduplnější nebo " sociálně A
51: Hudák. To, že *ekonomika* byla vlastně příliš <měkká> , nakonec musela přiznat i koalice ve A
55: lyrickým pasážím, které tolik vyhovují jeho <měkkému> , něžnému a civilnímu *projevu*. I os A
56: d of Paradise?... brumendem převzali kluci <měkký> , něžný *chorus*, jako hučení lesa... hey A
57: kterizována stupnicí : velmi měkká, měkká, <měkkčí> , normální, tvrdší, tvrdá, velmi tvrdá A
61: chtěl. Co je tvrdé, vzdorné, to se zlomí. Co je <měkké> , poddajné, to se ohne, ale nezlomí. C A
72: tři palce od jeho čenichu, a hovořila k němu <měkkým> , sípavým *pokuckáváním*, co chvíli A
81: na a jako Varvara dala tušit, že její sametově <měkký> , tmavý *mezzosoprán* neztratil nic ze s A
97: paláci. Galerie Velryba, jejíž problematicky " <měkká> " *koncepce* zahrnuje kvalitativně nev A
101: i na majitele Objevily se již spekulace, že " <měkký> " *postup* ČNB je motivován předvole A

Humans

- 5: sdržnost. To беру velmi vážně a nemíním být <měkký> . Ale na druhé straně se nemíním vy H
10: né s předváděním a přednáškou. Jsem *člověk* <měkký> . Pokaždé je mi prodávajícího líto, ž H
17: em si říkal, že letos se na to vykašlu, ale jsem <měkkéj> . Uvědomujete si, nakolik Lucie ovl H
44: ouchejte, slečno Meg ! Když ste v životě moc <měkká> , lidi vás využívají. To si pamatujte ! H
50: době mnohými viděn jako *člověk* zbožný, ale <měkký> , muž kompromisu. Arcibiskup Bera H
68: roto, že si myslím, že *lidé* zkrátka jsou takoví. <Měkkčí> , přízpusobiví, slabí, a proto chtiví, z H
71: musí být škvíra a ona mi ji ucpává. *Pepinka* je <měkká> , sametová, je moje. Cítím její lepka H
88: ávoji... Prokop měl oči plné slz ; cítil se slab a <měkký> , že se až styděl. Před šestou se však H
90: vel a ti jeho kamarádi nebyli tenkrát tak tuze <měkkčí> ! Kdyby ten Pithart neslyšel trávu rů H
92: ojena. Nejvíc jí vadilo, že manžel byl " takový <měkký> " . Otce popisovala jako autoritativ H
135: té). Nový hlavní konstruktér Mišín byl však <měkký> a nerozhodný. Projekt L - 1 nedok H
172: být právě tak bezmocný jako ten nesmělý a <měkký> *člověk*, jímž opovrhoval. Žena proh H

- 238: , maličká. Jsme tvrdé jako kámen a zároveň <měkké> jako dětská bačkora. Copak já vím, H
246: - AP Tvrdý obchodník z Dallasu je v jádru <měkký> KENNEDY BUDE MATKOOU V tém H
289: umí plést hebké svetry, v politice rozhodně <měkká> není, " soudí znalci, kteří bedlivě sle H
365: ční středisko pro lidská práva označuje jako <měkké> skinheady, dohlížely desítky polici H

Animal

- 69: se krmič opije a nepřijde. Některé krávy jsou <měkké> , pustí mléko samy, ale většinou m An
75: hnízdu, v němž seděla vrkající holubice, celá <měkká> , šedá, krásná - nádherný výtvar for An

Metaphors

- 7: či všem drogám, ani faktickou legalizací drog <měkkých> . Jenže právě toto " tvrdé jádro " o M
52: a lety byly v centru pozornosti policie drogy <měkké> , např. marihuana a hašiš, dnes už ve M
85: tomilí. Opravila jsem pak v duchu tvrdé y na <měkké> , uvědomujíc si je všechny tři. Ivana, M
96: lze však sotva očekávat, že by Dánové měli " <měkkí> " azylovou politiku než zbytek EU, p M
105: extem Bradleyho Strattona posluchačům s " <měkkíma> " ušima jako by tlumočí stoneovs M
141: pravdu zavřela brána. Za tím krajem, který je <měkký> a sladký jako tělo, a na čele hlavy M
151: Napsala omýtka s tvrdým y a dobili hrad s <měkkým> a zapoměl s ie a dokonce ve slově M
182: desetiletí odvážnou cestu uvolnění prodeji <měkkých> drog. Tento experiment přinesl ús M
183: á dohromady album na podporu legalizace <měkkých> drog. Účast zatím přislíbili mimo M
199: t na tvrdých drogách, jako je heroin. Přitom <měkké> drogy jako marihuana, jsou prý na s M
200: provázkem, drátem a podobně. Při vázání <měkkých> dřev dejte pozor, aby provázek ne M
267: Je třeba, abychom si vzájemně porozuměli. <Měkká> křídla evropského Fénixe Marcell v M
376: ního odběratele našich výrobků, které se na <měkkém> sovětském trhu nemusely příliš s M

Terms

- 211: láte ostrou špičku. Sklo podložíte plstí nebo <měkkým> dřevem a místo, kde má být díra, T
217: tává než pramínek ušlechtilosti nafilmované <měkkými> filtry, domnívá se list a píše o pr T
235: m v jeho testu se vyskytovali Přemyslovci s <měkkým> i. Zarážející jsou rovněž gramatick T
256: způsoby ukončení řádku je zásadní rozdíl. <Měkký> konec řádku dokáže editor při další T
257: jší akcí, nemůže dojít. Podobně jako tvrdý a <měkký> konec řádky, existuje i tvrdý a měk T
258: dý a měkký konec řádky, existuje i tvrdý a <měkký> konec stránky. Měkký konec stránk T
259: ky, existuje i tvrdý a měkký konec stránky. <Měkký> konec stránky vytváří editor podle z T
261: RÁK o šíří čelistí 60 až 80 mm s vložkami z <měkkého> kovu, protože bez něho nemohou T
262: lých hmot dát na čelisti vložky ve tvaru L z <měkkého> kovu (olova, hliníku), aby se pře T
263: jeden jemný na tvrdé kovy, jeden hrubší na <měkké> kovy, tvrdé umělé hmoty a dřevo v T

Idioms

- 377: u Manuelou nijak zvlášť nestál, " mám totiž <měkké> srdce, padre. " " Měkké srdce je do IF
378: doufat, že jste udělala dobře, ale máte příliš <měkké> srdce ! " Vtom lázeňské přivedly tři IF
379: jí). Ale na druhé straně mi teď došlo, že má <měkké> srdce (vždyť zatajila svou totožnost IF
380: nestál, " mám totiž měkké srdce, padre. " " <Měkké> srdce je dobrá věc, synu, ale v přípa IF
381: ana za roli ve snímku Frajer Luke, mužem s <měkkým> srdcem. Svědčí o tom i jeho vzta IF

Compiling a Monolingual Dictionary for Native Speakers*

Patrick Hanks, *Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic (patrick.w.hanks@gmail.com)*

Abstract: This article gives a survey of the main issues confronting the compilers of monolingual dictionaries in the age of the Internet. Among others, it discusses the relationship between a lexical database and a monolingual dictionary, the role of corpus evidence, historical principles in lexicography vs. synchronic principles, the instability of word meaning, the need for full vocabulary coverage, principles of definition writing, the role of dictionaries in society, and the need for dictionaries to give guidance on matters of disputed word usage. It concludes with some questions about the future of dictionary publishing.

Keywords: MONOLINGUAL DICTIONARIES, LEXICAL DATABASE, DICTIONARY STRUCTURE, WORD MEANING, MEANING CHANGE, USAGE, USAGE NOTES, HISTORICAL PRINCIPLES OF LEXICOGRAPHY, SYNCHRONIC PRINCIPLES OF LEXICOGRAPHY, REGISTER, SLANG, STANDARD ENGLISH, VOCABULARY COVERAGE, CONSISTENCY OF SETS, PHRASEOLOGY, SYNTAGMATIC PATTERNS, PROBLEMS OF COMPOSITIONALITY, LINGUISTIC PRESCRIPTIVISM, LEXICAL EVIDENCE

Opsomming: Die samestelling van 'n eentalige woordeboek vir moedertaalsprekers. Hierdie artikel gee 'n oorsig van die hoofkwessies waarmee die samestellers van eentalige woordeboeke in die eeu van die Internet te kampe het. Dit bespreek onder andere die verhouding tussen 'n leksikale databasis en 'n eentalige woordeboek, die rol van korpusgetuïenis, historiese beginsels vs. sinchroniese beginsels in die leksikografie, die onstabieleit van woordbetekenis, die noodsaak van 'n volledige woordeskatdekking, beginsels van die skryf van definisies, die rol van woordeboeke in die maatskappy, en die noodsaak vir woordeboeke om leiding te gee oor sake van betwiste woordgebruik. Dit sluit af met 'n aantal vrae oor die toekoms van die publikasie van woordeboeke.

Sleutelwoorde: EENTALIGE WOORDEBOEKE, LEKSIKALE DATABASIS, WOORDEBOEKSTRUKTUUR, WOORDBETEKENIS, BETEKENISVERANDERING, GEBRUIK, GEBRUIKSAANTEKENINGE, HISTORIESE BEGINSELS VAN DIE LEKSIKOGRAFIE, SINCHRONIESE BEGINSELS VAN DIE LEKSIKOGRAFIE, REGISTER, SLANG, STANDAARDENGELS, WOORDESKATDEKKING, KONSEKWENSIE VAN VERSAMELINGE, FRASEOLOGIE, SINTAGMATIESE PATRONE, PROBLEME VAN KOMPOSISIONALITEIT, LINGUISTIESE PRESKRIPTIVISME, LEKSIKALE GETUIENIS

* This article is an edited version of a plenary address delivered at the conference on 'Dictionaries, More Than Words', which took place at the Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia, 6 February 2009.

Introduction: dictionary and database

This article gives an account of the English experience in creating monolingual dictionaries aimed primarily at native speakers rather than foreign learners. It starts by comparing the role of a dictionary with that of a lexical database and saying a few words about the issues of register and correctness. Then, briefly, something will be said about words and word histories, lexicographic research, and coverage — what Dwight Bolinger called 'getting the words in'. I shall also discuss dictionary structure — both macrostructure and microstructure.

A lexical database is a fundamental background resource for use in the creation of many important linguistic artefacts — dictionaries, course books, computer programs for natural language processing among them. A great monolingual dictionary has a different function: it brings together speakers of a language, it has a socially integrative function, making explicit the basis of words and meanings and usage, which all uses of the language rely on. Words have meanings — or rather, strictly speaking, they have the *potential* to make meanings when put into context — and they are associated with particular sets of syntagmatic patterns, which can be discovered through painstaking corpus analysis. But words also have register: that is, not all words are equally appropriate in all circumstances. Some words and some grammatical structures are slang, or only used appropriately in spoken contexts, or characteristic of particular regions or dialects; others are only used in formal legal documents, or in romantic fiction or in poetry; others are meaningful and clear, but should not be used at all in polite society. These aspects are implicit in a database, on the basis of the kinds of texts in which each word and use occurs. But a good dictionary reports all of these aspects explicitly. It is not only an inventory of words, their meanings, and their syntagmatic patterns; it is also a report on matters such as register — social attitudes to 'correct usage'. The dictionary is expected to give rulings on what is correct and what is incorrect in different contexts in matters of usage. It is important that such rulings should be based on empirical analysis of the actual usage of good writers, rather than on the preferences and prejudices of a few journalists, academics, and self-appointed pundits, so there is a need here for interaction between a scientifically constructed lexical database and a dictionary as a social artefact.

Etymology and common sense

What is the role of a dictionary in researching and reporting etymology and word history? Some people think this is the *only* function of a dictionary, but it will be argued that an even more important function is the common sense one of identifying the conventions of word meaning and word use on which members of a language community rely in order to communicate with each other. A distinction must therefore be made between two major kinds of monolingual dictionary for native speakers. On the one hand, traditional major dictionaries

are based on historical principles and report word history and etymology. An example of this kind of dictionary is the multi-volume *Oxford English Dictionary* (OED). The first edition appeared in parts (called 'fascicles') between 1884 and 1928; the second edition was published in 1986; and the whole work is currently being revised at Oxford University Press under the editorship of John Simpson. The revision is available online, so that new research for each word is made available to the scholarly community and the public very soon after it has been completed. The new edition of OED is being made available piecemeal, as the project goes along, without constraints of alphabetical order. In the old days, we had to wait for up to fifty years for publication of lexical research in a big historical dictionary; now we get it within a few weeks of its completion. The OED is a dictionary on historical principles: it places the etymology of the word first and then gives the oldest known meaning of the word after the etymology. Recent developments come last. Word meaning is unstable — it changes quite rapidly — so this means that, in a historical dictionary, the current meaning of many words is placed last or nearly last and is preceded by one or more obsolete, obsolescent, or rare senses. So, for example, a dictionary on historical principles will tell you that a camera is a small vaulted room and next that it is the treasury of the papal curia. Somewhat later on it will tell you that a camera is a darkened room (a camera obscura) at the top of a house, with a hole in the roof, above which is a mirror. The camera obscura reflects images of the surrounding city or countryside on a light table in the room. The importance of this obscure term is that it is the link between our modern word *camera* and the historical meaning, 'small room'. Only right at the end of the entry does a dictionary on historical principles mention that a camera is an apparatus for taking photographs or for making movies.

The other kind of monolingual dictionary is a dictionary on synchronic principles. Basically, a synchronic dictionary reverses the order of senses, placing the modern meaning first. Thus, such a dictionary tells you first that a camera is an apparatus for taking photographs and making movies. It then goes on to explain where the word comes from and how the modern senses developed. The focus in a synchronic dictionary is on reporting conventional meanings and use, rather than on historical and etymological research. Examples of English dictionaries on synchronic principles include the *Encyclopedic World Dictionary* (1971) and *Collins English Dictionary* (1979), both designed and edited by Patrick Hanks, and the *New Oxford Dictionary of English* (1998), which Judy Pearsall and Patrick Hanks designed and edited and whose title was changed (by omitting the word *New*) for the second edition. In America, Houghton Mifflin publishes *The American Heritage Dictionary* (AHD 1969, now in its fourth edition). This owes its origin in the 1960s to the outrage felt by James Parton, publisher of *American Heritage* magazine, at the failure to deal with issues of register and correctness in Merriam Webster's *Third New International Dictionary*, unabridged (1961). Parton decided to commission his own dictionary, and AHD is the result. It is not the prescriptive work that Parton

was expecting, but it is a very good example of a dictionary on synchronic principles. In addition to definitions and examples, AHD contains many short articles on disputed or debatable points of word usage, in which the opinions of over 100 stylistic pundits are compared and collated. The essential principle of AHD is that it is a dictionary on synchronic principles.

The main Australian dictionary, *The Macquarie Dictionary* (1981, now in its fourth edition) is likewise a dictionary on synchronic principles.

What about one-volume dictionaries on historical principles? One might think that, since a one-volume dictionary is a practical tool, there would not be any, but in fact there have been several, and some still survive and thrive. A British example of a dictionary on historical principles was *Chambers Twentieth Century Dictionary*, which lasted for nearly a hundred years until in 1988 the publisher decided, for marketing reasons, to replace it with *Chambers English Dictionary*, and in 1993 by *The Chambers Dictionary*, a work presenting the language on synchronic principles. In 2010, after a long and distinguished history, that publisher went out of business.

In America, rather surprisingly, dictionaries on historical principles are still dominant, even among one-volume dictionaries. America's favourite dictionary — if sales are anything to go by — the *Merriam Webster Collegiate Dictionary*, is a dictionary on historical principles. This work was first published in 1898 and at the time of writing is in its 11th edition. It is very doubtful whether many if any of the purchasers and users of this work are aware of the fundamental difference in principles between this and a more commonsensical synchronic dictionary. It seems possible that some people may even read the dictionary and believe that the 'true meaning' of *camera* is the first one reported — a small room or the Vatican treasury — and that somehow an apparatus for taking photographs is merely an informal late development of low register, to be avoided by careful writers. That would not be an unreasonable inference, given the arrangement of senses, though of course quite wrong. It is not clear how a user of such a dictionary is supposed to divine the modern meaning of a polysemous word, if he/she does not know it already. Current changes in the business model for dictionary publishing raise the question whether there will ever be a 12th Merriam-Webster's Collegiate (or a Fourth Unabridged). Maybe these works will be superseded by continuously updated online versions. I shall return to this question towards the end of this article.

Some words have been stable in meaning since English began; others have changed their meaning more than once. Innumerable examples of lexical meaning change in English could be given. A sock originally meant 'a light shoe'; *dope* was originally a varnish, not a drug; *silly* formerly meant 'happy' and 'of low social class'. Why is the English word *magazine* evidently a cognate of French *magazin*, although they have very different meanings? In English, a magazine is a periodical publication or a part of a gun; in France, the word denotes a department store, where you go shopping. The unifying historical feature is that both go back to an Arabic word meaning 'storehouse'.

The arrangement and presentation of information such as this in a dictionary can be a critical and difficult undertaking. In modern English *size* means bigness, dimension, magnitude — a very fundamental concept. One might imagine that the word *size* has always been part of English, but it has not. It is actually a late medieval development — a rather surprising one — due to the cheating habits of medieval bakers. In the 15th century, if you didn't like the loaf your baker gave you — if you thought it was too small, and if the baker consistently gave small measures — he might be taken to the local assizes, a court of law, and punished for unfair trading. So a size loaf was a loaf that was of a dimension or magnitude approved by the court. And from that narrow basis the word broadened outward to mean the dimension or magnitude of anything. The point is that meaning change in words is unpredictable, but it is a common and substantial feature of linguistic development. Today's exploitation of a word sense may become tomorrow's norm.

It was mentioned earlier that word meaning is unstable and may change often. Some people — including bilingual lexicographers — deny the existence of word meaning altogether. And you can see why: if you think of the word *fire*, what does it mean? Is it something burning out there in the field or the forest, out of control; or is it something burning in your house and nicely under control? Has it something to do with guns? Or has it something to do with losing your job or with making pottery? Perhaps it means inspiring enthusiasm? The answer is that *fire* in isolation means all of those things and much more as well — or rather, it has the potential to have these meanings, when used. For reasons such as this, it can be argued that, strictly speaking, a term in isolation has only meaning potential, not meaning. What you get in a monolingual dictionary is a list of meaning potentials, not of meanings. You need context to know which meaning is activated when a word is used; and in order to know what the normal contexts of words are corpora and corpus analysis is needed. As a source of data and a research technique, introspection does not work. We have learned the hard way, through fifty years of generative linguistics, that introspection does not provide reliable data about how words are really employed in everyday usage. One of the most important findings of corpus linguistics is that people, even trained linguists, are not very good at reporting their own linguistic behaviour. Introspection distorts, perhaps because people consulting their intuitions tend to think up unusual examples — illustrating the boundaries of possibility, rather than normal everyday usage. Normal usage seems to be buried so deep in the subconscious that it is hard for people to recall it to the conscious mind and report it accurately.

Getting the words in

The first duty of a lexicographer is to get the words in. The editor of a dictionary for native speakers must aim at a very wide inclusion policy, for very often it is the rare and unusual words and senses that people will want to look up.

The editor of a dictionary for foreign learners, on the other hand, will aim to be more selective, presenting and explaining just those words that, in his/her judgement, a foreign learner will need to know.

In 1857, Richard Chenevix Trench, one of the founders of the *Oxford English Dictionary*, described lexicographers as 'the inventory clerks of the language'. This seems exactly right. Compiling inventories may seem a lowly occupation, but in fact, as far as the lexicon is concerned, it is a task full of interesting challenges. It may seem that compiling an inventory of all the words in a language and saying what they mean should be easy, but in fact it is hard, it is difficult. Introspection does not yield an inventory of all the words in one's language, partly because of the difficulty of recall, and partly because nobody knows everything. And then there are problem cases, such as deciding what counts as a word. The core vocabulary of the standard language shades outwards in many directions — into technical jargon, regional dialect, slang, archaic vocabulary, poetic coinages, and so on. Should a dictionary include proper names? Where should a lexicographer draw the line? And when all those decisions have been made, there remain questions about word boundaries. English is not an agglutinating language like Turkish or — some would say — German, where the problems of finding word boundaries are much more serious, but nevertheless deciding on word boundaries in English raises a couple of interesting questions. Just two examples will be discussed.

The problem of phrasal verbs in English is well known. It seems obvious that *take off* — what a plane does when it leaves the ground and starts to fly — is not the same word as the base verb *take*. It has been said that *take off* has as about as much to do with *take* as *disease* has to do with *ease*. Clearly, then, these two verbs should be treated as separate lexical items in the dictionary. But then should a dictionary aim to include all phrasal verbs? What about phrasal verbs such as *finish up*? Should that be an entry in the dictionary too? The standard answer is no, because the meaning is compositional: in this expression, *finish* still means 'finish', and the force of *up* is merely completive-intensive. But then there are phrasal verbs such as *break up* which have both idiosyncratic and completive-intensive meanings:

A gathering such as a political demonstration can break up, or can be broken up by the police; a married couple or two people in a relationship can break up (in which case they are no longer in a relationship); in American English, when people break up, they are overwhelmed by emotion and start laughing or crying; and there are several other meanings for which both the verb *break* and the particle *up* must be present. But then there are meanings where the particle is optional: for example, you can break a table, or you can break a table up. Here, the role of the particle is to reinforce the meaning of *break*, suggesting that the table gets broken into not just two but several pieces.

In a dictionary, should all meanings for such verbs be given, or only the idiosyncratic ones?

Even more problematic are multiword expressions such as *fire engine* and

fire extinguisher. A fire engine is a truck carrying equipment for putting out fires. The meaning here is clearly not compositional: *engine* does not mean 'truck'. So *fire engine* should be an entry in the dictionary. The American term, *fire truck*, is more compositional. The meaning of *fire extinguisher* is likewise more or less compositional — it is a piece of equipment for extinguishing fires. But because this is a term that denotes a class of artefacts, each of which is a unique object, *fire extinguisher* is usually selected as a dictionary entry. Now, what about other multiword expressions such as *wood fire* and *forest fire*? There are thousands of such multiword expressions in English, and it is a highly productive area of the language: new multiword terms are being coined all the time. Dictionaries do not do a very good job of reporting them. The usual justification for omitting them is that their meaning is compositional: a wood fire is a fire burning wood; a forest fire is a fire burning a forest. But, although this may seem plausible at first sight, ultimately it is not satisfactory. *Wood* and *forest* are synonyms, so if the meaning were truly compositional, *wood fire* and *forest fire* ought to be synonyms too. But they are not. A wood fire is burning wood (a mass noun) under human control in a hearth in the home or in a camp — but a forest fire is burning a forest or forests (a count noun); it is raging out of control in the wild. The two terms have been conventionalized in different ways, employing different meanings of their basic components. Strictly speaking, a dictionary should explain this. But none do.

Accompanying policy decisions about what counts as a lexical item and where to draw the line is the question, how and where to find the words? During the past 150 years or so, the Oxford Reading Programme has been devoted to reading texts and collecting citations for the words used in them. During its heyday in the late 19th and early 20th century it involved many volunteer readers.

So, before corpora, hundreds of volunteers contributed citations on which the OED was based. Citation collection is a good way of collecting data, especially data for rare and unusual words. But readers have to exercise judgement in deciding what to collect. No one sends in citations for all the uses in a text of ordinary words such as *come* and *go* or *up* and *down*. So, although it was and is a wonderful and admirable enterprise, the Oxford Reading Programme necessarily introduced selective distortion. It did not provide reliable statistical evidence of usage. James Murray, the founding editor of OED, showed that he realized this when he complained in his presidential address to the Philological Society in 1878 (a time when he had only just started on the monumental task of sorting out the citations and writing dictionary entries on the basis of their evidence), 'We have fifty citations for *abusion*, but less than five for *abuse*.'

Citation readers collect citations for unusual words like *triskaidekaphobia* 'irrational fear of the number 13' and for unusual senses. Computers, on the other hand, do not exercise judgement. If asked to find all occurrences of the word *of* in a text or corpus of texts, a computer will do so in a few milliseconds. It can put them in a concordance, otherwise known as a KWIC ('key word in

context') index, which a lexicographer can sample and study and use to compare the patterns associated with each key word. Murray would have seen the point instantly of a large electronic corpus.

Another source of words is existing dictionaries. Lexicography is accretive. One dictionary builds on top of another dictionary. We do not all start from scratch. Dictionary writers are sometimes accused of plagiarizing each other's work, but if you think about it, every dictionary definition can be seen as a small hypothesis. No scientist would publish a hypothesis without consulting the work of his/her predecessors. So it is reasonable to look at other dictionaries and evaluate what they say. What is not reasonable, of course, is mindless copying. That's a danger, for (unlike any other form of research) lexicography requires the lexicographer to say something about everything. Copying is a temptation for the lexicographer. But it is a danger that, in a reputable project, must be resisted. Evaluating definitions in existing dictionaries in the light of new evidence is one thing. Copying out those definitions mindlessly is another.

Thus, there are three main sources of words for a dictionary: citation reading, existing dictionaries, and corpus evidence. Searching corpus data electronically has so far provided only a low yield for new words and new senses, partly because of the difficulty of deciding what counts as a word or sense, and partly because any corpus is only a sample of the language. Searching the Internet may eventually prove more productive, if ways can be found of rigorously defining the existing inventory and defining what counts as a 'lexical item'. As we have seen, identifying uniquely meaningful multiword expressions is a task that poses particular problems. One technique that has been proposed is measuring statistically significant changes in frequency of words and collocations. Thus, twenty years ago, in the infancy of corpus linguistics, Ken Church and Patrick Hanks were able to measure a sudden increase in the frequency of the word *greenhouse* and to note a new pair of associated multiword expressions: *greenhouse effect* and *greenhouse gas*. These terms are now standard entries in monolingual dictionaries.

Guidance on usage

The needs of users of dictionaries must now be looked at more closely. It has been said that a dictionary has a socially integrating function. This is true up to a point, but only up to a point. There are some good studies of the use of bilingual and foreign learners' dictionaries, but there are no good studies of dictionary use among native speakers, so what follows will necessarily be rather anecdotal, based in part on feedback from marketing departments at Collins in the 1970s and 1980s and at Oxford in the 1990s. Dictionary publishers — who in many cases control lexicographic budgets — often insist that a new dictionary should be 'market driven'. This is rather dangerous, a recipe for extreme conservatism, because until a product has been created, the public — the po-

tential market — has no way of knowing whether it will want the new product or not.

In English, disappointingly for lexicographers who work so hard on definitions and grammar, it seems that people use dictionaries mainly for spelling. This is almost certainly true, at least of dictionaries of the English language, the spelling of which is not phonetic but contains many irregularities and idiosyncrasies. On the other hand, maybe inflections (morphology) and dialect differences are less of a problem in English. Dictionary makers must offer guidance where guidance is needed.

People look to a dictionary for guidance, not only on spelling and inflections, but also on correct usage and word choice. Should we say 'uninterested' or 'disinterested', is there a difference? Nowadays you can hear people saying, 'I totally refute that'. No, you cannot refute a proposition by declaration. You can say, 'I deny that', because *deny* is a performative verb, like *promise*: you can deny or promise something merely by saying so. *Refute*, however, was not a performative verb until recently. To refute a proposition, in the traditional meaning of the word, effective argumentation is needed, not just performance of a speech act. Now, however, it is being used as a strong synonym of *deny*. 'I totally refute that' is politicians' speak for 'I don't want to acknowledge that it's true'. The dictionary should explain that careful writers and speakers still make a distinction in meaning between the two words.

Another example of the sort of guidance that a dictionary should give concerns grammatical complementation. Should one say 'bored with' or 'bored of'? More and more people say 'I'm bored of that'. Is it right or wrong? The usage of an increasingly large number of educated speakers of English cannot be ignored. There is no logical argument against *bored of*. Prepositional choice represents a set of arbitrary conventions. This example contrasts with other common usages, which can be objected to on logical grounds: for example, 'He could of done it' and 'He should of done it', which, though common, are errors. Here, the auxiliary verb *have* is clearly required; it has been replaced by the preposition *of*, which in rapid speech is a homophone of *have*, as a result of grammatical ignorance.

Likewise, 'between you and I', which is even more common, is objectionable on the logical grounds that English prepositions govern object-case pronouns (in no variety of standard English does anyone say, 'He gave it to I' or 'She came home with I'); there is a perfectly good object-case pronoun, *me*. This error is a result of the death of grammatical case in English (except for a few pronouns) coupled with hypercorrection. What has happened is this: schoolchildren are taught that it is wrong to say 'me and' in subject position (as in 'Me and my friends are going on holiday'), but they do not fully understand the nature of the grammatical error, so, with hypercorrection, they use 'and I' in place of 'me and' on all occasions, regardless of case. Thus, 'Between you and I' has become established as a formula among people who have no sensitivity to grammatical case. The question is, should a dictionary acknowledge such errors as part of English, and if so, what should it say?

Another example concerns the so-called split infinitive. In English, for at least three hundred years, there has been a lively debate — especially among people who believe that English is really Latin in disguise — about whether it is acceptable to 'split' an infinitive by putting an adverb between the infinitive marker and the verb. Can you say 'to boldly go [where no man has been before]'? There is no logical objection to this, but conservative self-appointed pundits object to it. The dictionary should give a ruling.

Lexical and paralexical content

It is fashionable in some academic circles to make a distinction between 'lexical semantics' and 'encyclopedic information'. But ordinary monolingual dictionary users do not make the same careful distinction. People want instant cultural reference information and do not care whether it is classified as 'semantic' or 'encyclopedic'. Here are examples of the sort of questions in English that people ask and expect to have answered by a dictionary:

- 'What's the scientific name for a thrush?'
- 'Is your scapula your shoulder blade, your backbone, or your collarbone?'
- 'What's the capital of Chile?'
- 'Why is a madrigal called a madrigal?'
- 'What does *nook-shotten* mean?' (We find Shakespeare talking about England as a nook-shotten island. What does it mean?)
- 'What is a predator?' people might ask. 'Is a penguin a predator?' [Well, they catch fish, don't they?]
- 'What are chinos?'
- 'What's an ohm? What's a joule, and why is it called a joule?'
- 'Is *aa* an English word?'

People use their dictionaries for Scrabble and for crossword puzzles. As a matter of fact, *aa* is an English word: it is a kind of volcanic lava, a word of Hawaiian origin — not very common in everyday reading and conversation, but remarkably useful for Scrabble.

In addition, people want to have a dictionary as an authoritative inventory of their language, even if the dictionary sits on their bookshelves and they never look inside it. They want it there on the bookshelf just in case they might one day want to look something up.

If they do look inside a dictionary, sometimes people just want to browse. So they also want fun words. For example, here are some words denoting criminals of various kinds from different periods and different sources: a cutpurse (a street thief; used by Shakespeare), a mosstrooper (mosstroopers were criminal raiders on the borders between England and Scotland in the 16th and 17th centuries). What's a yegg? What's a snakehead? What's a tsotsi? What's a

rudeboy? What's a grifter? These are all different kinds of criminals in different parts of the English-speaking world. And above all, the marketing department will tell you, 'We want *new* words, because then the journalists will write about our dictionary.'

Corpus evidence and examples

A corpus shows how each word is used. It does not show directly what each word means, but it provides evidence on the basis of which meanings can be inferred. The first editions of the two large monolingual dictionaries of British English (Hamlyn 1971 and Collins 1979) were designed and edited before corpus evidence became available. Then, in 1983, in the earliest days of corpus lexicography, John Sinclair and others started working on the first edition of the COBUILD dictionary. They discovered that many of the generalizations made in pre-corpus dictionaries, though plausible, were not quite right. That is, they did not stand up well to comparison with corpus evidence. The *New Oxford Dictionary of English* (1998), mentioned earlier, was the first (and, so far, the only) dictionary for native speakers of English to be based on corpus evidence and well as citations from a reading programme.

Corpus evidence provides an essential source of information for collocations and syntagmatics, which need to be studied statistically in order to understand the relationship between word use and word meaning. This provides a structure or framework of a dictionary. Patterns of word use can be detected in corpora, but these patterns provide hints, associations, and probabilities about meaning and usage, rather than certainties. They point, in fact, beyond lexicography to a need for new lexically based approaches to linguistic theory.

An essential design feature of a natural language is that it is full of uncertainty. This is because the categories found in natural languages are built around prototypical 'best examples' and have boundaries that are fuzzy, rather than being sharply defined. This, of course, presents a big problem for lexicographers. To take a very simple example, corpus evidence shows that the prototypical use of the verb *hazard* in English is with the noun *guess* as a direct object. The phraseology is 'hazard a guess' in more than 50% of the uses of this verb in all the corpora consulted. The meaning of the phrase as a whole is, 'to say something without much confidence that it is true'. The prototypical direct object, *guess*, is a noun denoting a speech act or a concept. On closer inspection of the corpus (for example, the British National Corpus), all sort of other speech-act and concept nouns are found in the same slot — not only the near synonym *conjecture*, but also *inference*, *opinion*, and even *definition*. There are even some examples of the verb governing indirect and direct speech, as in 1 and 2. These are boundary cases, ungrammatical in most varieties of American English.

1. I would **hazard** that the ratio of real balances to total private sector net worth is less than 1% ...
2. "My uncle," said Wendy, expanding further on her family, "was Provost of Dumfries; he had a rather odd name — 'Chicken'." "Not Hen Chicken?" I **haz-arded**, as this humorous diminutive was part of my family mythology.

The dilemma for the lexicographer in such cases is whether to represent and gloss the prototypical example (in this case, *hazard a guess*) or whether to set such a broad scope (*hazard something*) that the normal phraseology and its meaning are in danger of being lost sight of. There is no single correct solution to this dilemma. It is a matter of judgement and choice, taking account of the likely needs of the intended users. A useful compromise involves using the word 'typically', for example by defining the first sense of *hazard* as 'to state a proposition, typically a guess, without any great confidence that it is true'.

If a broader scope is chosen for the definition of this sense, typical phraseology can be highlighted by means of an example. For this and other reasons, the dictionary maker should resist the temptation to choose weird, inventive, creative, unusual boundary-case examples, and instead choose examples that represent central and typical, normal usage, even though such usage may seem slightly boring. The objective in selecting examples should be to illustrate normal usage, not to illustrate the boundaries of all imaginable possibilities. Unfortunately, for some reason there is a strong human urge to focus on boundary cases and unusual usage, so young lexicographers have to be *trained* to select examples that are normal, even boring.

Interpreting the evidence

The example of *hazard*, verb, shows how corpus evidence can augment and even supplant intuitions. The first thought of many English speakers, consulting their intuitions without the benefit of objective evidence, is that this verb means 'to put at risk'. It certainly does have this meaning, but only in about 20% of all uses, if the corpus evidence is to be believed. The corpus nudges the lexicographer into recognizing facts of the language that are not intuitively obvious.

On the other hand, an example of how *not* to use evidence and examples in a dictionary may be given from Wiktionary; in the monolingual English version, the verb *hazard* is defined as:

1. To expose to chance; to take a risk: *I'll hazard a guess.*
2. To incur or venture.

These definitions appear to have been copied, with minor alterations, from another dictionary, with no thought as to how the word is actually used. An example has been tacked on to sense 1, although it actually illustrates sense 2. This is not an isolated error; indeed, it is fairly typical of the monolingual English Wiktionary.

There is a dramatic contrast between Wikipedia and Wiktionary. Wikipedia has been a great success. It is a vast anthology of encyclopedic articles written by people claiming to have knowledge about a particular subject. If that claim turns out to be ill-founded, i.e. if an article turns out to be erroneous, then, if it is of any public interest at all, it is pounced on by genuinely knowledgeable people and improved or replaced. It seems that the model of an encyclopedia as a collectively written anthology is a good one. This model, however, cannot be extended to a dictionary. A dictionary is not an anthology. Wiktionary is full of second-hand derivative entries, often wrongly defined or with erroneous examples. For reliable information about the words and meanings of a language, the evidence of actual usage — corpus or citations — must be interpreted by knowledgeable and trained people following a set of consistent principles.

Writing definitions

As indicated in the preceding section, the first priority for a monolingual lexicographer is to give shape to each dictionary entry, writing definitions that reflect the evidence by selecting a middle course somewhere between accounting only for prototypical uses of a word and accounting for all imaginable uses.

The next priority is to achieve technical accuracy in definitions. Writing definitions of technical terms is a particular problem for monolingual lexicographers. In order to understand and explain a term — and definitions should aim to *explain*, not merely to define — the definer needs to be a user of the terms being explained. This applies not only to scientific definitions but also to other domains such as sports. Anybody who has ever played cricket knows how badly worded American dictionary definitions of cricketing terms can be. For example, in one recent American dictionary, we are told that the bowler in cricket is 'the player who throws the ball to the batsman' — there is no mention of the obligatory straight arm that distinguishes bowling from throwing by fielders and from pitching in baseball. The distinction is important because the sporting pages of English-language newspapers outside North America quite often contain sentences such as:

Brett Lee, Australia's answer to Shoaib Akhtar, is the latest fast bowler to be accused of throwing. — Simon Briggs, *Daily Telegraph*, London, 22 February, 2009.

A reader trying to interpret this sentence and consulting the American definition of *bowler* just mentioned, would be puzzled rather than enlightened.

No doubt Americans find the same kind of bad wording in British definitions of baseball terms, or American football, which they call football, or hockey, which we call ice hockey. Definers need to be users of the terms being defined to appreciate the importance of technically accurate components of

meaning. But expert users of terms who are not trained lexicographers are often particularly bad at defining them and explaining them, so there needs to be interaction between the technical adviser — the scientist or the sportsman, as the case may be — and the lexicographer, who has the skill of defining and succinctly explaining.

In some words there is a clash between the meanings used by the scientific community and the meanings of ordinary people. Ordinary people, when they say 'wait a second', do not mean 'wait the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom'. This, however, is the definition of *second* as the basic scientific unit of time, agreed by the General Conference on Weights and Measures, which meets from time to time in Paris, as part of the *Système International d'Unités* (SI units). A serious dictionary must give both the scientific definition and explain the ordinary language usage.

A similar problem arises in defining many everyday creatures and other objects, e.g. *spider*. Consider the following extract from the *Oxford Dictionary of English*:

an eight-legged predatory arachnid with an unsegmented body consisting of a fused head and thorax and a rounded abdomen. Spiders have fangs which inject poison into their prey, and most kinds spin webs in which to capture insects. Order Araneae, class Arachnida.

The first part of this entry aims to define — to set boundaries around a classification — rather than to explain. Why mention 'an unsegmented body' and 'a fused head and thorax'? These features are mentioned in order to distinguish spiders from insects, which form a completely different zoological class. It is the second sentence in this entry — which is not part of the formal definition — that goes some way towards explaining. Now contrast this with the COBUILD entry:

A spider is a small creature with eight legs that looks like an insect. Most types of spider make webs in which they catch insects for food.

This entry is clearly more concerned with explaining matters to foreign learners than with scientific definition. Definitions in monolingual dictionaries aim to do both.

Using a corpus such as the British National Corpus and with the help of a corpus analysis tool such as the Sketch Engine, it is possible to compile a corpus-based linguistic profile of terms such as *spider*.

- Many thousands of species of spiders are known (*funnel-web, web-building, orb-weaving, bird-eating, ground-dwelling, giant, huge, large, tiny, poisonous, black widow, camel, redback, trapdoor, wolf, whitetail, crab, tarantula*, etc.).
- Some species of spiders *hunt* prey.

- Spiders *bite*.
- Some species of spiders are *poisonous*.
- Many species of spiders *spin webs*, with threads of *strong silk*.
- Spiders *lurk* in the centre of their *webs*.
- Spiders *control* what is going on in their *webs*.
- Spiders have eight *legs*.
- Their legs are *thin*, *hairy*, and long in proportion to body size.
- Spiders have *eight eyes*.
- Spiders spend a lot of time being *motionless*.
- Spiders' *movement* is *sudden*.
- Spiders *crawl*.
- Spiders *scuttle*.
- Spiders are *swift* and *agile*.
- Spiders can *run up walls*.
- Many people have a *dread* (*hate, fear*) of spiders.
- People *kill* spiders.
- English people are much concerned with trying to get spiders out of the *bath*.

Such a profile summarizes the beliefs that most people, at least in England, have about spiders. (With regard to the last point in this list, it should be added that, although there is a mildly significant association between the lexical items *spider* and *bath* in some British corpora, it is not suggested that this is a serious fact requiring scientific investigation.)

Consistency of sets

Once the framework for each ordinary word has been created using corpus evidence, other kinds of information must be slotted in. The principle of coverage of terms in all fields of human activity, including sports, leads to another principle of monolingual lexicography, namely consistency of sets. All the terms in a set — the chemical elements, for example, or the organs of the human body — should be defined in a similar and consistent style, regardless of frequency, i.e. even though some members of the set may be so rare that they do not show up at all in a corpus. The same principle holds good for the terminology of activities such as snooker, curling, and Australian rules football — all of which have sprung into prominence only in recent years, disseminated mainly by television. If people watching snooker on TV hear the commentator say, 'There is a possible plant here', they may well turn to the dictionary for a relevant definition of *plant*. *The Oxford Dictionary of English* defines this as 'a shot in which the cue ball is made to strike one of two touching or nearly touching balls with the result that the second is potted'. This in turn implies that there must be an adequate definition of the snooker sense of *cue ball* and the verb *to pot*.

Thus, when a lexicographer reads a newspaper or watches TV, it is often the case that the lexicographer is less interested in the content of what is being said than in how it is being said. What are the words being used? Do they need to be dictionary entries?

The editor of a monolingual dictionary has to decide how far to go in technical fields. Should *strobilus*, *strobila*, *strobilation* be entries? Native speakers who do not know these words may expect to find them in a dictionary. On the other hand, a dictionary is not a termbank. The terminology of the sciences in particular but also of technological activities is so vast and specialized that much of it does not belong in a dictionary, but rather in a project like IATE (Inter-Active Terminology for Europe), the 23-language terminology database of the European Union, which collects and stores technical terminology, much of it of an extremely abstruse variety, with stipulative definitions.

What all this adds up to is that native-speaker dictionary users expect the inventory of words in a dictionary to be complete, and the lexicographer must find ways of satisfying that expectation, despite the fact that the goal is impossible: the lexicon of a living language is dynamic and the boundaries of its vocabulary are fuzzy, so that new words and expressions are being coined — invented or borrowed — all the time.

New words

Most dictionary publishers issue, with each new edition of a major dictionary, a booklet of new words to excite the journalists. The *Macmillan English Dictionary for Advanced Learners* (MEDAL) is no exception. It is actually a dictionary for foreign learners, but the publishers well understood the virtues of publicity, so in 2008 they issued a small, free booklet that included such 'new words' as *blogosphere*, *chav*, *air kiss*, *career gapper* (somebody who's taken time out from their career), *Chelsea tractor* (one of those tank-like vehicles supposedly driven especially by people who live in Chelsea, a somewhat rich district of London). *Chick lit* is a genre of literature written for young women to read, typically while sunning themselves on the beach; *civil partnership* usually denotes a homosexual marriage. A *designer baby* — it is amazing what you can do with genetics these days.

The need to get on with it

Compiling an authoritative monolingual dictionary for native speakers is a daunting task. A natural language consists of thousands of lexical items, ranging from extremely common items such as function words and light verbs to rare technical terminology and compounds. The editor-in-chief must develop clear policies for all of the issues mentioned in this article, together with many others, and ensure that they are followed consistently by other contributors

working together, not merely as a team, but as a 'single collective author'. The whole policy cannot be established before actual entry writing begins. Rather, broad outlines are established at the outset, which are then extended and modified as the project goes along, in response to particular issues that come up.

Conscientious definition writers tend to agonize over capturing the precise meaning of each word. But agonizing is counter-productive. It sometimes happens that the first form of words that a lexicographer jots down is a perfectly good one; they then agonize and gradually make the entry less and less satisfactory and less and less comprehensible, typically by trying to cover all eventualities. Fear of making a mistake is another factor that slows lexicographers down without bringing any noticeable benefit. The sad fact is that slow writers who agonize make just as many mistakes as quick writers working according to a good plan. The plan needs to be outlined very broadly at first, then developed as the project starts up. The details cannot be developed satisfactorily in advance, in an abstract theoretical vacuum.

This problem is best dealt with by setting up a system for a dictionary project where each compiler is free to do his/her honest best and move quickly on. The system says, 'Don't worry about making a mistake; somebody else will check what you have written.' The editor-in-chief will check and ensure that obvious errors and accidental infelicities are corrected, and give feedback. Lexicography is a team game; the team should have a structure; lexicographers should read and check each other's work in a co-operative environment.

The medium

In the modern world, the question arises, in what medium will any new monolingual dictionary be published? The traditional medium of a bound book for reference information is being superseded by the Internet. This raises major questions for the future of lexicography, including the following:

- Is the Internet as secure and durable a medium as the printed page? Will future readers, in five hundred years time, be able to consult an electronic dictionary on the Internet in the same way that a present-day reader can consult an old book in a library?
- Can a new dictionary any longer be a product created within the capitalist system as an investment by a publisher or risk capitalist? Or must all new dictionaries be funded by central, government-controlled funding agencies, as they were in the former totalitarian states of Eastern Europe?
- If funding is to come from a commercial investor such as a publisher, what is the business model? People have got into the habit of expecting information to be freely available and contributed by volunteers, on the model of Wikipedia, but, as discussed above, this is not a satisfactory

model for a dictionary. Is it realistic to expect sufficient revenue to accrue from advertising to justify the huge investment required to fund the creation of a good new dictionary, or can the habits and expectations of online users be changed, so that they will pay for the information they obtain?

Conclusion: evidence and interpretation

A large modern monolingual dictionary of any language has an important role to play in the community, and language communities, large and small, need their dictionaries. An essential requirement is a lexical database, constructed by analysis of corpus evidence, but also reflecting social attitudes to language. A corpus shows patterns of word usage. Supplementary research is also needed for terminology, unusual words, names, word histories, attitudes to correctness, and other matters — but such research is all designed to find evidence, not to promote the opinions of self-appointed pundits. At the core of lexicography, therefore, lies the corpus. The basic task is to report all normal uses and meanings of all normal words. But a dictionary must also reflect social attitudes to language and give guidance on meanings and etymology, but to be authoritative, all pronouncements must be based on evidence of usage — corpus evidence. Public attitudes to points of 'correct' usage should be reported and evaluated. If a dictionary tries to cover all imaginable possibilities of use of any content word in the language, there is a danger of lapsing into incoherence. The language will defeat the over-ambitious lexicographer. This is because word meaning and use is infinitely flexible. What a dictionary is reaching for is the central norm that speakers of a language rely on when they speak to each other, not the wild imaginings of linguists dreaming up remote possibilities.

Acknowledgement

This work was funded in part by the Academy of Sciences of the Czech Republic (project T100300419) and the Czech Ministry of Education (National Research Program II project 2C06009) as part of a corpus-driven investigation of lexical issues at the Institute of Formal and Applied Linguistics of the Charles University in Prague.

References

Dictionaries

- Delbridge, A. et al. (Eds.). 1981. *The Macquarie Dictionary*. Sydney: The Macquarie Library.
- Gove, P.B. (Ed.). 1961. *Webster's Third New International Dictionary of the English Language*. Unabridged. Springfield, Mass.: Merriam Webster.

- Hanks, P. (Ed.).** 1971. *Encyclopedic World Dictionary*. London: Hamlyn.
- Hanks, P. (Ed.).** 1979. *Collins Dictionary of the English Language*. London: Collins.
- Morris, W. (Ed.).** 1969. *The American Heritage Dictionary of the English Language*. Boston: Houghton Mifflin.
- Mish, F. (Ed.).** 2004. *Webster's Eleventh Collegiate Dictionary*. Springfield, Mass.: Merriam Webster.
- Pearsall, J. and P. Hanks (Eds.).** 1998. *New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Rundell, M. and G. Fox (Eds.).** 2002. *Macmillan English Dictionary for Advanced Learners*. London: Macmillan.
- Sinclair, J. (Ed.-in-Chief).** 1987. *Collins COBUILD English Language Dictionary*. London/Glasgow: Collins.

Other Literature

- Bolinger, Dwight.** 1970. Getting the "Words" In. *American Speech* 45(1-2): 78-84.
- Trench, Richard Chenevix.** 1857. On Some Deficiencies in Our English Dictionaries. Hartmann, R.R.K. (Ed.). 2003. *Lexicography: Critical Concepts*: 171-216. London/New York: Routledge.

National Language and Terminology Policies — A South African Perspective

Mariëtta Alberts, *Department of Language Management and Language Practice, University of the Free State, Bloemfontein, South Africa*
(malberts@lantic.net)

Abstract: Terminology plays a pivotal role in language development and the promotion of multilingualism. This article discusses the issue of multilingualism regarding terminology policies as seen from an African perspective. Special emphasis is given to the South African situation regarding language policies and consequential terminology policies in terms of subject-oriented terminography, translation-oriented terminography and linguistic community-oriented terminography.

Keywords: LANGUAGE POLICY, LINGUISTIC COMMUNITY-ORIENTED TERMINOGRAPHY, LINGUISTIC DIVERSITY, MULTILINGUALISM, SUBJECT-ORIENTED TERMINOGRAPHY, TERMINOLOGY DEVELOPMENT, TERMINOLOGY MANAGEMENT, TRANSLATION-ORIENTED TERMINOGRAPHY

Opsomming: Nasionale taal- en terminologiebeleid — 'n Suid-Afrikaanse perspektief. Terminologie speel 'n sleutelrol in taalontwikkeling en die bevordering van veeltaligheid. Hierdie artikel bespreek die kwessie van veeltaligheid ten opsigte van terminologiebeleid soos gesien vanuit 'n Afrikaperspektief. Spesiale klem word gelê op die Suid-Afrikaanse situasie rakende taalbeleid en die gevolglike terminologiebeleid in terme van vakgeïntereerde terminografie, vertaalgeïntereerde terminografie en taalgemeenskapgeïntereerde terminografie.

Sleutelwoorde: MEERTALIGHEID, TAALBELEID, TAALDIVERSITEIT, TAALGEMEENSKAPGEÏNTEERDE TERMINOGRAFIE, TERMINOLOGIEBESTUUR, TERMINOLOGIEONTWIKKELING, VAKGEÏNTEERDE TERMINOGRAFIE, VERTAALGEÏNTEERDE TERMINOGRAFIE

Unless people who speak African languages realise the importance of using them, the status of their languages will not improve.

Nomso Mgijima, *The Sowetan*, 23 February 2007

1. Introduction

Information is distributed and knowledge is acquired through terminology. The terminology of each subject field or domain is increasing with every new development or invention. The supplying of appropriate scientific, technical, educational and economic terms should be a national priority, especially in a multilingual dispensation.

Should the terminology of the minority/marginalised/developing or standardised languages of the country be developed into functional terminologies, South Africans would be equipped with effective communication tools. Terminology development is also a vehicle appreciating the innovative skills of the language and subject-related communities within the country. In this sense, subject areas such as science, technology and economy can play a role in the development of languages as they have done with English, French, German, Chinese, Japanese, and Afrikaans to mention just a few. Languages can develop into functional languages through efforts of terminology development by language offices, private initiatives and publishers.

Terminology plays a pivotal role in language development and the promotion of multilingualism. The availability of (multilingual) polythematic terminology is an indicator of development since specialised communication has a central axle or hub in terminology. Standardised terminology contributes to the quality of translations, editing, interpreting services, dictionary compilation and specialised or subject related communication. Streamlined translation and interpreting services provide competitive advantages.

It is important to develop official languages into functional languages in all spheres of life. Information transfer, assimilation and retrieval should be through the first language or mother tongue. It is proven that information is best acquired (decoding process) and conveyed (encoding process) through the first language. Standardised terminology leads to exact communication and misinterpretation or misunderstanding are avoided.

Terminology, therefore, is a strategic resource and has an important role in a country regarding the functional development of languages and their users. Effective economic, scientific and technological transfer and assimilation of knowledge and skills amongst subject specialists and laypeople, and the communication skills of the citizens of a country are developed through the use of correct terminology. Although terminological and terminographical activities are not always cost-effective, they are of invaluable cultural, social, historical, functional, academic and scientific importance.

Terminology as a discipline, however, is governed by the language policy of the country. It is at the service of the language policy of the reigning government, e.g. monolingual, bilingual or multilingual. The South African situation could be taken as example:

- previous dispensation (i.e. prior to 1994): bilingual technical dictionaries were compiled (English/Afrikaans);
- present dispensation (i.e. after 1994): multilingual term lists are compiled in the eleven official languages of the country.

The concept "multilingualism" is very complex and the meaning or connotation attached to it in Europe may differ from that in Africa: "Africa has over 2 000 languages and a rich and diverse linguistic heritage" (Chin 2009). In Africa,

many different languages are spoken within the same political and geographical areas. The minimum number of languages spoken in African countries such as Nigeria, Ghana, Angola, Mozambique, Gabon, etc. is 40 for each country. Nigeria has in addition to the dominant languages such as English, Hausa, Ibo, Fulani and Yoruba more than 200 other languages spoken in the country. The Democratic Republic of Congo has more than 250 languages including the major languages, i.e. French, Swahili, Lingala, Kikongo and Ciluba. It is difficult, if not impossible, to be conversant in all these languages. Governments therefore need to recognise standard varieties spoken by majority groups as official languages. Languages foreign to Africa such as English, French, Portuguese, German, Spanish and Italian are not only spoken in almost 90% of the African continent, but these languages are also used as lingua franca between the indigenous African communities. Besides being used as lingua franca, these colonial languages are also used as official languages over and above the indigenous African languages. Some of the official indigenous languages are completely neglected and left to become extinct. Africans were to a certain extent made to believe that it is civilised to speak a "prestigious" European language. Reasons for this preference could be ascribed to aspects such as jealousies among Africans (i.e. not being prepared to accept another language to be elevated to official status); prestige or status of languages (standard vs. dialect); the colonial language is available, functional and developed; economic incentives; international status; and a wider audience for communication purposes. These attitudes regarding preference for colonial languages lead to the extinction of indigenous languages since children are deprived of learning and using their linguistic heritage (cf. Mojela 2007).

2. Language policies of South Africa

Wright (2007: 6) says about language planning and language policy:

Language planning and language policy express particular efforts at official social intervention and control in the sphere of language. Normally the direction and ambitions of language policy and planning embody an economic and social vision. In accord with this vision, certain aspects of current language practice in society are officially challenged or curbed, others are sustained and affirmed. In general, the aim of language policy is to move language practice in directions deemed desirable by those in power. Usually such attempts are applied through legislative measures ('policy') and allied material provision ('planning') to different social and political entities, such as geo-political regions, organised economic alliances, nations, provinces, industries, school systems, government departments, businesses and so forth.

In 1964 Valter Tauli (cf. Wright 2007: 6) defined language planning as "the activity of regulating and improving existing languages or creating new common regional, national or international languages. In 1992 Christopher Brumfit

(cf. Wright 2007: 6) described language planning as "the attempt to control the use, status, and structure of a language through a language policy developed by a government or other authority". The *Random House Dictionary of the English Language* concurs, but adds some significant detail: Language planning is "the development of policies or programmes designed to direct or change language use, as through the establishment of an official language, the standardization or modernization of a language, or the development or alteration of a writing system".

According to Wright (2007: 7), a distinction can be made between natural and interventional language planning. Natural language planning actively supports the evolving language needs of a society as they emerge in response to other-than-linguistic pressures. Interventional language planning is prepared to challenge the impact on the language dispensation of current sociolinguistic forces. It would revitalise declining languages, preserve dialects, maintain languages that are under threat, modernise traditional languages for use in different domains, defend language rights, and nurture an ethically satisfying linguistic ecology. Existing language policies reflect different combinations of these two approaches.

The language policy of a country influences terminology development. Such a policy determines the number of languages to be developed as functional languages:

- Eastern Africa: One official language (Swahili), various national languages
- Namibia: Since 1992, only one official language (English), 16 national languages
- South Africa: Prior to 1994, two official languages (English and Afrikaans). Since 1994, eleven official languages (with special emphasis also on the development of South African Sign Language (SASL) and Khoe and San languages), many dialects.

Section 6 of Chapter 1 of the Constitution of the Republic of South Africa, 1996 (Act 108 of 1996) declared Sesotho sa Leboa (Northern Sotho), Setswana (Western Sotho or Tswana), Sesotho (South Sotho), Tshivenda (Venda), Xitsonga (Tsonga), Afrikaans, English, Siswati (Swati), IsiNdebele (South Ndebele), IsiXhosa (Xhosa) and IsiZulu (Zulu) as official languages of South Africa. According to the Constitution, "the state must take practical and positive measures to elevate the status and advance the use of these languages".

The South African language policy promotes the equitable use of the official languages. It also accommodates linguistic diversity. The language policy provides a regulatory framework to facilitate the effective implementation of the constitutional obligations concerning multilingualism.

In accordance with the Constitution and the National Language Policy Framework (NLPF) and Implementation Plan (2003), the Government aims to promote South Africa's linguistic diversity. It proposes that optimal use be made of the country's linguistic resources by fostering respect for linguistic diversity and, by implication, for linguistic rights. Government aims to achieve this by means of an approach of functional multilingualism. This implies that the choice of a particular language in a particular situation is determined by the context in which it is used, i.e. the function, the audience and the message for which it is employed. The main criteria of functional multilingualism are language preference, use and proficiency.

Should the South African Government therefore need to communicate with the citizens of the country it is stipulated in the National Language Policy Framework (2003) that Government documents "shall be made available in all 11 official languages". In cases where this is not feasible, National Government Departments "shall publish documents simultaneously in at least six (6) official languages", i.e. Tshivenda, Xitsonga, Afrikaans, English, one from the Nguni group (i.e. IsiNdebele, Siswati, IsiXhosa, IsiZulu), and one from the Sotho group (i.e. Sesotho sa Leboa, Setswana, Sesotho). The three Sotho languages are mutually intelligible as are the four Nguni languages.

In occupations which are directly language-based, such as translation, interpreting, lexicography, terminography and copy-editing, and occupations which are directly language-dependent, such as in teaching, the media and tourism, the economic role of language is self-evident. Language is, however, also of central importance in the economic life of a country, i.e. as the major means of human communication, training (e.g. vocational training), effective management, trade negotiations, provision of services, job security and institutional loyalty and the job market (cf. Webb n.d.).

Considering the role of language in economic activity, it is a given that language can be either a barrier or a facilitator to economic activity. The role of language in economic performance is conditioned by the linguistic character of the community. In countries like Germany, France and the United Kingdom, which are largely dominated by one language, and where by far the majority of the population know that language, language may be less obviously an economic issue. In multilingual or highly multilingual countries, however, like practically all African states, language is exceptionally central, as is apparent from the situation in South Africa (cf. Webb n.d.).

Although the linguistic diversity and language services are costly, National Treasury, nonetheless, concluded after thorough investigation that it is possible to implement a multilingual language policy in South Africa. The expenditure is relatively small compared to overall departmental budgets. The result to aim for: enhanced communication between Government and the citizens of the country.

The distribution of the official languages in various provinces is an indication of the language variety in the country (PanSALB 2001a; see Table 1):

| Home language | Eastern Cape | Free State | Gauteng | KwaZulu-Natal | Limpopo | Mpumalanga | Northern Cape | North West | Western Cape | South Africa |
|------------------|--------------|------------|---------|---------------|---------|------------|---------------|------------|--------------|--------------|
| Afrikaans | 9,3 | 11,9 | 14,4 | 1,5 | 2,3 | 6,2 | 68,0 | 7,5 | 55,3 | 13,3 |
| English | 3,6 | 1,2 | 12,5 | 13,6 | 0,5 | 1,7 | 2,5 | 1,2 | 19,3 | 8,2 |
| IsiNdebele | 0,1 | 0,4 | 1,9 | 0,2 | 1,5 | 12,1 | 0,1 | 1,3 | 0,0 | 1,6 |
| IsiXhosa | 83,4 | 9,1 | 7,6 | 2,3 | 0,3 | 1,5 | 6,2 | 5,8 | 23,7 | 17,6 |
| IsiZulu | 0,8 | 5,1 | 21,5 | 80,9 | 0,7 | 26,4 | 0,3 | 2,5 | 0,2 | 23,8 |
| Sesotho sa Leboa | 0,0 | 0,3 | 10,7 | 0,1 | 52,1 | 10,8 | 0,1 | 4,2 | 0,0 | 9,4 |
| Sesotho | 2,4 | 64,4 | 13,1 | 0,7 | 1,3 | 3,7 | 1,1 | 5,7 | 0,7 | 7,9 |
| Setswana | 0,0 | 6,8 | 8,4 | 0,1 | 1,6 | 2,7 | 20,8 | 65,4 | 0,1 | 8,2 |
| Siswati | 0,1 | 0,3 | 1,4 | 0,1 | 1,1 | 30,8 | 0,1 | 0,6 | 0,0 | 2,7 |
| Tshivenda | 0,0 | 0,1 | 1,7 | 0,0 | 15,9 | 0,2 | 0,0 | 0,5 | 0,0 | 2,3 |
| Xitsonga | 0,0 | 0,3 | 5,7 | 0,0 | 22,4 | 3,8 | 0,0 | 4,7 | 0,0 | 4,4 |
| Other | 0,2 | 0,2 | 1,0 | 0,4 | 0,3 | 0,3 | 0,7 | 0,6 | 0,4 | 0,5 |
| Total | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |

Table 1: Distribution of the official South African languages in various provinces

The Constitutional principle (limiting factor) governing the choice of languages to be used in official sectors of society is found in section 6(3)(a) of the Constitution. This section stipulates that usage, practicality, expense, regional circumstances and the balance of the needs and preferences of the population as a whole or in the province concerned (i.e. demographic, economic and attitudinal factors) will be taken into account in choosing the languages to be used in a specific sector (cf. Constitution 1996; NLPF 2003; PanSALB 2001b).

From a language planning point of view, the languages which people other than the educated elite understand and in which they are most competent, are as follows in descending order: IsiZulu, IsiXhosa, Afrikaans, and Setswana.

English is hardly used for neighbourhood communication in the Limpopo, Mpumalanga, North West, Free State, Northern Cape and Eastern Cape. Its functional use in neighbourhoods is limited to KwaZulu-Natal, Gauteng and the Western Cape. English is therefore useful as a lingua franca in only three provinces. IsiZulu is similarly used as a neighbourhood language in three provinces: KwaZulu-Natal, Gauteng and Mpumalanga. The use of Afrikaans as a neighbourhood language extends across five provinces: Northern Cape, Western Cape, Free State, Gauteng and Eastern Cape. As such it has a wider spread than any of the other languages (cf. PanSALB 2001a).

It is clear that there is no realistic opportunity for a single lingua franca across South Africa. A combination of Afrikaans and IsiZulu would take care of seven provinces. Limpopo and the North West have language needs for neighbourhood communication which are very specific to their respective provinces and which do not include English. IsiZulu ranks third after Afrikaans and English as a language of the educated elite. English is barely known amongst people with a less than Grade 10 qualification (cf. PanSALB 2001a).

One of the factors that ameliorates the asymmetrical treatment of languages in South Africa is the considerable multilingualism that is evident. This is reflected in the following results (PanSALB, 2001a; see Table 2):

| | | |
|---------------------------|-------------------------|-----|
| Afrikaans speakers | English | 50% |
| | African languages | 11% |
| English speakers | Afrikaans | 54% |
| | African languages | 14% |
| Sesotho speakers | Afrikaans | 24% |
| | English | 28% |
| | Other African languages | 30% |
| Setswana speakers | Afrikaans | 24% |
| | English | 14% |
| | Other African languages | 20% |
| Sesotho sa Leboa speakers | Afrikaans | 7% |
| | English | 19% |
| | Other African languages | 13% |

| | | |
|---------------------|-------------------------|-----|
| Siswati speakers | Afrikaans | 6% |
| | English | 27% |
| | Other African languages | 42% |
| IsiNdebele speakers | Afrikaans | 10% |
| | English | 3% |
| | Other African languages | 19% |
| IsiXhosa speakers | Afrikaans | 7% |
| | English | 24% |
| | Other African languages | 13% |
| IsiZulu speakers | Afrikaans | 4% |
| | English | 32% |
| | Other African languages | 16% |
| Tshivenda speakers | Afrikaans | 4% |
| | English | 0% |
| | Other African languages | 13% |
| Xitsonga speakers | Afrikaans | 2% |
| | English | 24% |
| | Other African languages | 21% |

Table 2: Levels of understanding of languages other than their own among home language groups

These results seem to illustrate the fallacy of assuming that a lingua franca exists in South Africa. Clearly South Africans need a more inclusive language policy than one based on the assumption that one language has sufficient reach to be an adequate medium of communication across the country (PanSALB 2001a).

Several varieties of informal languages exist around South Africa. These include the varieties known as Tsotsitaal/Flaaitaal and Isicamtho, which are patois developed from a blend of languages and serving as bridges for communication. They even differ from region to region in South Africa depending on the languages prevalent in a particular area (PanSALB 2001a).

South Africans are surprisingly generous in attempting to accommodate the practical challenges that arise in a country with eleven official languages. Substantial portions of citizens, however, have to accept less than adequate communication with their fellow South Africans and have to make difficult adjustments in formal situations and economic interaction.

Among better educated South Africans there is a clear indication of principled commitment to their own languages and the conviction that more should be done to develop minority languages. Among the less educated and rural communities there are a similar level of frustration based not on principles and ideals of linguistic equality, but on the reality that fluency in languages other than their own is critical for sheer economic survival (PanSALB 2001a).

3. Language development

The term *language development* is defined in various ways. Haugen (1966) defines it as a process of changing a selected variety from an oral language to a written one. Cluver (1996: 1) reasons that language development is not a one-time process, but rather a continuing process referring to the planned modification of a selected variety to fulfil any new function that it did not fulfil previously.

Language development should be part of and contribute to the overall development of a community. One of the objectives of language development is to enhance the status of the language with its own speakers and with foreign speakers. The status of a language could be enhanced by proving to its users that it can be employed as a modern means of communication to function in domains in which languages of wider distribution function, i.e. education, the legal system, local administration, health care and modern agriculture (cf. Cluver 1996: 1-2; 6). Language development should refer to the whole language and include moves to develop the literary part of language as well (Cluver 1996: 7).

According to Cluver (1996: 2-3; 6), language development entails:

- *language selection* — the selection of a specific variety (e.g. dialect) for development,
- *language codification* — the development of a writing system, documentation of the existing vocabulary and grammar, development of spelling and orthography (for an unwritten language) or modernising existing spelling and orthographies and standardising the language,
- *language elaboration* — the expansion of the functions of the language so that it can operate beyond its traditional domains (home, family, community) as working language in the public domains (government offices, court of law, etc.),
- *language spread* — the development of the language as a language for teaching and learning (domain of education for primary school level, secondary school level and tertiary level), and a language of communication (e.g. media), and
- *language modernisation* — general terminology development in all spheres of the working environment, e.g. economy, science and technology (i.e. the expression of abstract concepts).

Budin (2004: 6) alleges:

Language development is much more than coining words and terms and thinking up spelling reforms. It is an ecological approach to language as a crucial element in human societies. It is more than language planning and standardization. It also includes multiple socio-linguistic factors.

He further maintains that language development covers a mix of methods and approaches, including terminology and lexicography, terminology management, translation work and translation management, and increasingly corpus-based approaches (term extraction, corpus analysis for identifying neologisms coined in discourse communities, etc.).

Cluver (1996) claims that language welfare should reflect the welfare of the speech community and that language development should form an integral part of the development of any community by the community itself. He pleads for the first objective for language development in South Africa to be the production of a standard grammar, a standard general dictionary and technical terminologies for each of the official languages.

The practical problems of accommodating eleven official languages are such that current budgets cannot provide for adequate expenditure on language development, multilingual training and the creation of translation and interpretation services. There is a danger in a situation in which resources will always be limited that the challenges of language development will always be deferred. Government and decision-makers in the public and private sectors, however, know that one of their major obligations is to be understood by the citizens of the country. If less than half of the population is reached because of communication problems, there should be enough reason for a greater sense of urgency in respect of language development and multilingualism.

Currently various governmental and private initiatives are undertaken, aimed at promoting multilingualism by the redressing of past linguistic imbalances, developing the previously marginalised languages, and supplying multilingual polythematic terminologies.

4. Brief historical overview of South African terminology practice

The history of African language terminology in South Africa starts with structures similar to the erstwhile Language Boards that begun in 1928 with the formation of language committees (cf. Mayevu 1996: 30). These have been changing over the decades both in formation and focus (cf. Mtintsilana and Morris 1988).

Since 1948 the construction of the Language Boards has taken a fundamentally political outlook both in its composition, function and relation. These Boards were linked to 'independent homelands' and similar political formations. Their political role was to present a picture of separate development, primarily the homeland structures. Some of these Boards were actually accountable to the Chief Ministers and their Cabinets, some were accountable to the Department of Education and Training and some to both structures. These Boards were mainly funded by the homeland governments and had representatives of those governments as members (cf. Mtintsilana and Morris 1988). There were Language Boards for every African language, though some would be duplicated for political purposes as in the case of Transkei and Ciskei and

Bophuthatswana (cf. Ntshangase 1996).

Officially the South African terminology practice started as early as 1950. Before then many terminology lists or technical dictionaries were compiled by individuals. The early stages of the official terminology practice were adhering to the then bilingual policy of the country.

Translators of the erstwhile Language Bureau of the Department of Culture (later Department of National Education), started documenting English and Afrikaans terms on index cards. Later a Terminology Division developed within the Language Bureau. Terminologists were appointed to excerpt terms from documents and to systematise and standardise the terminology of various subject fields. A variety of bilingual term lists and technical dictionaries were compiled and published.

Other language bureaux situated at government departments followed soon (e.g. the South African Defence Force (SADF), South African Railways and Harbours (SAR&H), Department of Education, etc.). Soon thereafter several language bureaux at institutions such as the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) (South African Academy for Science and the Arts), South African Iron and Steel Corporation (IsCOR), Municipalities, South African Bureau of Standards (SABS), South African Broadcasting Corporation (SABC), etc. that employed translators also had a few terminologists to excerpt the terminology of the relevant institution. They all started to compile bilingual terminology lists in order to standardise the terminology usage within these organisations. A Coordinating Terminology Board (COTERM) was formed in 1971 to coordinate the terminology endeavours of the various offices and to avoid duplication. The members of COTERM also started with initiatives to computerise the terminologies of the various organisations.

In 1976 the Terminology Division of the Department of National Education amalgamated with the Vaktaalburo (Terminology Bureau) of the SAAWK to form a new Terminology Bureau under the auspices of the Department. This Bureau was later named the National Terminology Services (NTS) of the Department of Arts, Culture, Science and Technology (DACST) and the office became the official national terminology office to document and disseminate terminology. This office took over the duties of COTERM.

In April 1998 the NTS and the State Language Services of DACST amalgamated to form the National Language Service (NLS). The NTS became the Terminology Coordination Section (TCS) of the NLS under the Department of Arts and Culture (DAC).

Although the TCS and its predecessors in the past primarily compiled bilingual, bi-directional technical dictionaries, it was soon realised that there was a pressing need for terminology in the various African languages. The TCS therefore started with research into word forming principles in the indigenous languages. It was decided to start compiling multilingual technical dictionaries in all national South African languages with the addition of extended information such as definitions and explanations.

In 1994 South Africa obtained a multilingual policy and since then terminology projects are undertaken in the eleven official languages. African language terminologists, who are all first-language speakers of the various official languages, were employed since 1995 to document African language terminology on a variety of subject fields.

The TCS of the NLS provides a base of knowledge for all language groups on national, provincial and local government level and for all other spheres of technical and scientific communication. The office serves the whole community. This incorporates all levels of communication in all registers in which terminology of the official languages is needed — from grass-roots level to the higher echelons of science and technology.

The TCS of the NLS operates in close collaboration with external official and private multilingual projects, and with the Pan South African Language Board (PanSALB).

5. Establishment of the Pan South African Language Board (PanSALB)

The Pan South African Language Board (PanSALB) was established in 1996 to give effect to the letter and spirit of Section 6 of the Constitution of the Republic of South Africa. PanSALB is a constitutional body instituted in terms of the PanSALB Act (Act No 59 of 1995 as amended in 1999).

The Board was established to promote multilingualism and develop the official South African languages, including the South African Sign Language (SASL) and the Khoe and San languages.

PanSALB created advisory structures to assist it in achieving its mandate, namely to promote multilingualism, to develop languages, and to protect language rights.

PanSALB structures consist of

- nine Provincial Language Committees (PLCs) to assist the provinces with language policy formulation and implementation.
- thirteen National Language Bodies (NLBs) to take care of standardisation (e.g. spelling and orthography rules), terminology development, dictionary needs (general and specific), literature and media, research, and education.
- eleven National Lexicography Units (NLUs) to compile comprehensive monolingual and other types of dictionaries (i.e. bilingual translation dictionaries, etymological dictionaries, technical dictionaries, etc.).

Members of the PanSALB structures have a major communicative role to play in terms of information flow in all three directions. They should also monitor PanSALB's strategic objectives, implementation plan and the effectiveness thereof.

The PLCs are provincial structures, each taking care of the languages of a specific Province and assist with the language policies of these provinces. The PLCs keep in contact with the NLBs and NLUs in order to make them aware of all the language needs that impact negatively or positively in terms of language policy, language practice, language implementation, and the promotion of multilingualism in a province.

The NLUs and NLBs are national structures. The language-specific NLB and NLU take care of the particular language or language group, the official language (where applicable) of the majority of the speakers of the language residing in the geolinguistic area, developmental issues regarding the language, and promotion and preservation of the language.

The NLUs were established as Section 21 Companies. Although they receive funding from PanSALB, they are autonomous. They are managed by Editors-in-Chief and function under the auspices of Boards of Directors. Their main aim is to compile comprehensive monolingual explanatory dictionaries to preserve and document the respective official languages. They may, according to needs, also compile bilingual and other types of dictionaries.

PanSALB's NLBs function according to technical committees (TCs) to develop standards, spelling and orthography rules, conduct research, verify and authenticate terminologies, and assist with the standardisation of terms, determine dictionary needs (for general and special purposes), facilitate production and promotion of literature and media, develop dialects, facilitate research studies, develop all bands of education, and facilitate the development and promotion of the SASL and the Khoe and San languages.

The TCs concerned with Terminology Development obtain terminology lists from the Terminology Coordination Section (TCS), the National Language Service (NLS), the Department of Arts and Culture (DAC), and other institutions or individuals who compile technical dictionaries or term lists. Their members could form part of the working groups who are busy compiling ongoing terminology projects. They could assist with research on problematic terms, term creation or word-forming principles relevant to the language concerned. They also give feedback to the parties concerned.

Once a terminology project is finalised the project is submitted to the relevant Technical Committee (TC): Terminology Development of the relevant NLB to verify and authenticate terms, assist with the standardisation and stabilisation of terms, popularise terms, advise compilers of technical dictionaries or term lists (i.e. private or institutionalised efforts), advise terminologists and terminology offices, and co-operate, collaborate and communicate with stakeholders.

The TC: Terminology Development should not only give feedback regarding term creation to the compilers of the terminology project, but also to the NLB concerned. They should furthermore give feedback to various institutions busying themselves with term creation regarding specific language developmental needs, requirements of the specific language and subject-oriented

needs. These institutions should on the other hand, change their databases according to the discussions held with NLB members on, for instance, problematic terms. TC members could give input regarding term creation, spelling and orthography, word-forming principles, principles regarding neologisms, borrowing, transliteration, total embedding, etc.

The TC: Terminology Development of the NLB should present the term list to the NLB concerned to give its stamp of approval concerning the terminology project before the term list or dictionary could be disseminated to end-users (e.g. subject specialists, language practitioners, NLUs, and language users in general).

PanSALB drafted guidelines to assist each technical committee with the various tasks to be handled.

6. Terminology policy and official terminology management in South Africa

The need for unambiguous communication in theoretical and applied fields of human activity is constantly increasing. Unfortunately people use terminology without realising what it entails. It is the process of excerption, documentation, standardisation, publication and dissemination of terms. This is a time-consuming and labour-intensive process where specialised skills, commitment and devotion are needed.

A terminologist aims to provide unambiguous terms for well-defined concepts. In order to achieve this aim, it is necessary to determine the precise meanings of terms to enable users to comprehend and use them in a universally accepted manner.

Since the beginning of terminology development in South Africa, the source language (SL) for terminology documentation is usually English. The Afrikaans terminologies for various subject areas and domains were officially documented since 1950. Afrikaans terminologists and subject specialists therefore created terms for almost all subject areas and domains. This, however, is not the case for the other nine official South African languages. Owing to the great backlog concerning the development of these terminologies, various methodologies are being used to harvest terminology, i.e. subject-oriented terminography, translation-oriented terminography, and linguistic community-oriented terminography.

6.1 Subject-oriented terminography

The official South African terminology office, the Terminology Coordination Section (TCS) usually starts from a specific source language (SL), namely English, and translation equivalents have to be supplied in the other ten official South African languages. The national terminology office uses various methods

but usually works according to the subject-oriented methodology for terminology harvesting.

The terminologists consult linguists and subject specialists when supplying or coining translation equivalents for SL terms. The National Terminology Services (NTS) used to work in close collaboration with the old Language Boards and at present the terminologists of the National Language Service (NLS) consult the National Language Bodies (NLBs) established by PanSALB.

The terminologies of all official languages should be developed to enhance the multilingual heritage of the country and to develop the languages into functional languages in all subject areas and domains. Unfortunately there are very few trained terminologists with even less terminology posts available.

The national terminology office, the Terminology Coordination Section (TCS), manages terminology as follows:

- The terminographers excerpt terminology in the SL which is usually (but not necessarily) English.
- The terms are then supplied with definitions, example sentences and relevant information in the SL.
- These terms and relevant information are discussed with subject specialists to confirm the contents.
- After the SL terms and relevant information have been finalised, the information is converted/translated into the target languages (TLs), e.g. Afrikaans and the nine official African languages.
- The information in the target languages is also discussed with collaborators and subject specialists before finalisation of the dictionary.

Language-specific collaborators, situated in the various provinces, were trained by senior TCS staff members on the basic principles and practice of terminology work. The idea is that these collaborators (mostly linguists or language practitioners) form the core group to assist the language-specific terminologists. When the terminologists for instance compile a Mathematics term list, the core group of collaborators would get subject specialists working in the field (e.g. teachers or lecturers in Mathematics) who are first-language speakers of the relevant language to assist them with the discussion of the terms and related information. The terminologist would then return to the office and change the database according to the suggestions made by the collaborators and subject specialists.

After the finalisation of a specific terminology list, the relevant NLB is requested by the TCS to verify and authenticate the terminology. After verification and authentication, the relevant terminologist changes the database according to the suggestions made by the language-specific NLB members.

After the database has been finalised (e.g. all eleven official languages have been taken care of), the relevant terminology list can be published and disseminated to target users (cf. Diagram 1).

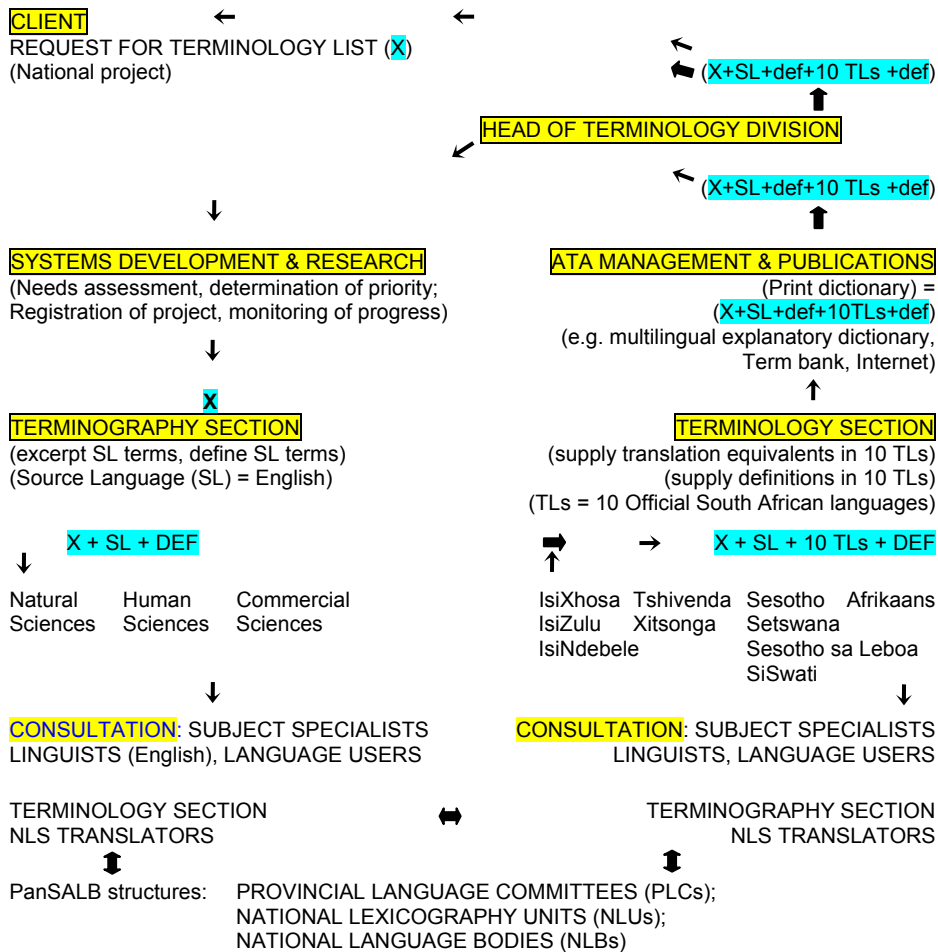


Diagram 1: Terminology flow diagram

6.2 Translation-oriented terminography

The National Language Service (NLS) also employs translators and therefore also takes part in translation-oriented terminography (TOT). Terms in the source language (SL) text and term equivalents in the target language (TL) translations are aligned and harvested (cf. Diagram 2). These terms and related information are then submitted to the Terminology Coordination Section (TCS)

to be documented and processed in the manner described above and illustrated in Diagram 1.

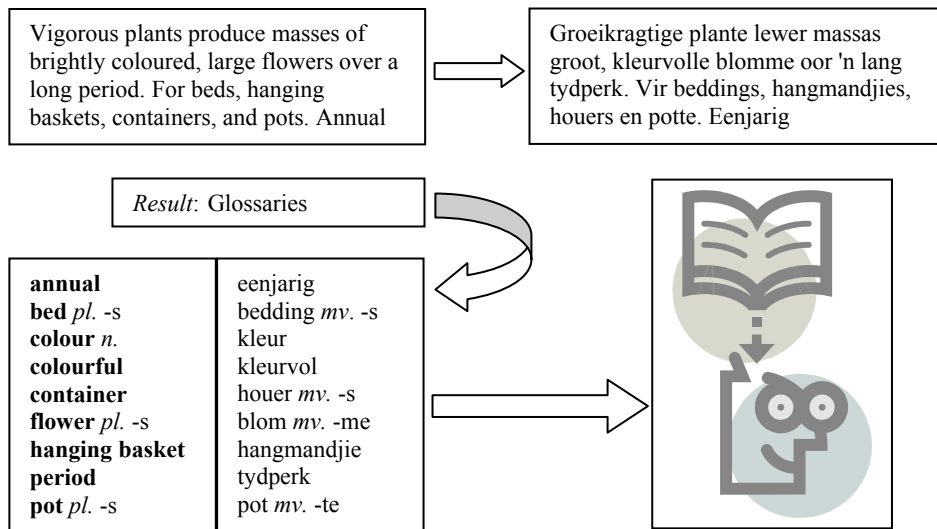


Diagram 2: Translation-oriented terminography

6.3 Linguistic community-oriented terminography

Terminology is also harvested from rural and urban speech communities (community-oriented terminography) for documentation in the central terminology bank. This process which entails field-work is costly and therefore seldom undertaken. The fact that terms existing in the communities are not documented contributes to the general stereotype that the African languages are incapable of naming abstract concepts. This is a fallacy. Various terms in a variety of domains already exist in the African languages. The problem is that these terms are not documented and therefore not standardised. The various dialects also contain a wealth of terms which could be harvested and utilised in the standard languages.

For terminology development in a specific language to be effective, trained terminologists should be in a position to do field-work to obtain and document the terminology that exists in various subject areas and domains in the relevant language. Since the terms are not documented in a systematic format (e.g. database), they cannot be standardised. More than one term for the same concept exist in the same language, leading to duplication, confusion and poor communication. The language-specific terminologist should visit rural areas to document the terminology related to animal names, bird names, customs and beliefs, traditional medicines, etc. (i.e. linguistic community-oriented terminography). These terms could be obtained from the older members

of the linguistic community. Should these people die, the knowledge of terms and related information (i.e. indigenous knowledge systems) dies with them.

The language-specific terminologist should also do field-work in the urban areas — especially at language offices and tertiary institutions where the terminologist could excerpt terms and related information from translated documents, textbooks, curricula, master's dissertations and doctoral theses at various departments (e.g. Physics, Zoology, Chemistry, Psychology, Sociology, Art, etc). These terms are new, and if they are not documented (and translated into the relevant official languages where needed), they could not be standardised and disseminated to target users. It is only by documenting terms and related information, standardising terms and disseminating term lists in various domains and subject areas that the South African indigenous languages will become functional languages in all spheres of life.

This argument also underpins the principle of translation-oriented terminology where translators harvest the source language terms and their target language equivalents when translating. The National Language Service (NLS) established pilot Language Research and Development Centres (LRDCs) in geolinguistic areas and, while in existence, these Centres also functioned as depositories of harvested terminology which was then submitted to the Terminology Coordination Section (TCS).

Should PanSALB employ trained terminologists at the various National Lexicography Units (NLU), these terminologists will be situated at tertiary institutions where they could harvest terminology as described above. They are already situated in the geolinguistic area where the most first-language speakers reside and would be in a position to do field-work as suggested. They will also be able to contribute multilingual polythematic terms to the envisaged Human Language Technologies (HLT) virtual network (see section 9).

PanSALB and the National Language Bodies (NLBs) would be in a better position to assist with terminology development in the country. Terminology endeavours by various language-specific compilers of technical dictionaries could be coordinated. The terminology of a specific language could be verified by the relevant NLBs the moment the term list is compiled. The specific language could publish the terminological information immediately after the verification and authentication process and after applicable changes were made to the database. The terminology list could be disseminated to target users such as language practitioners, journalists, etc. without having to wait for the other languages to be completed (such is the case for national projects). All these terms and related information should be submitted to the national term bank at the NLS.

7. Terminology models in Africa to be implemented by South Africa

South Africa is the only country in Africa where terminology is officially practised separately from lexicography. The South African model functions well in

cases where national terminology projects need to be undertaken (e.g. school terminology lists for the national Department of Education). Unfortunately this model does not work for the general terminology development of a specific language.

The models of respectively the Institute for Kiswahili Research at the University of Dar es Salaam, Tanzania and the African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe could serve as examples where both lexicography and terminography work are done at the same institution.

In the case of the Institute for Kiswahili Research, the lexicographers work on a variety of general dictionaries, e.g. monolingual Swahili dictionaries (explanatory dictionaries; school or learner's dictionaries, etc.) or various types of bilingual dictionaries (Swahili–English). It also houses terminologists who compile monolingual or bilingual technical dictionaries in various subject fields, e.g. Physics, Chemistry, Biology, etc.).

In the case of ALRI, the lexicographers compile monolingual, bilingual and multilingual dictionaries. Monolingual general dictionaries for Shona and Ndebele have already been published (e.g. *Duramazwi reChiShona* (1996), *Isichazamazwi SesiNdebele* (2001); *Duramazwi Guru reChiShona* (2001), etc.). These dictionaries are only some of several subprojects undertaken by the African Languages Lexical (ALLEX) Project at the University of Zimbabwe. At the same time, the terminologists at ALRI have compiled a number of technical dictionaries in Shona and Ndebele (sometimes in combination with English), e.g. *Duramazwi reUrapi neUtano* (2004), *Duramazwi reMimhanzi* (2005), *Isichazamazwi SezoMculo* (2006), *Duramazwi reDudziramutauro neUvaranomwe* (2007), etc.

In South Africa, official terminology work could still continue on a national basis by the Terminology Coordination Section (TCS) as described in section 6. Language-specific terminologists could assist with the documentation of terminology that at present exists in the various indigenous languages (e.g. created by translators, interpreters, journalists, the community radio, the media, and the linguistic community at large).

Computerised versions of the terminology documented by language-specific terminologists could be sent to the national terminology office, TCS, for inclusion in the national term bank managed by the NLS. This work could be done online since all National Lexicography Units (NLUs) and other terminology compilers work on TshwaneLex and TshwaneTerm (locally developed custom-made software) which are compatible with MultiTerm and TRADOS.

The terminological inputs to the national term bank by collaborators such as translators working in various government departments, national parliament, the erstwhile Language Research and Development Centres (LRDCs), PanSALB structures such as the NLUs and National Language Bodies (NLBs), and private terminology initiatives could solve the problem of term documentation and dissemination. Present problems experienced in South Africa regarding the process of standardisation and communication could be solved.

8. Terminology training

The Terminology Coordination Section (TCS) of the National Language Service (NLS), Department of Arts and Culture (DAC) provides in-house training as official national terminology office. Unfortunately there is only one terminologist per language at TCS. These terminologists are not enough to solve the subject-related communication problems since they cannot cope with the demand to supply multilingual terminology in the diversity of subjects or domains that need multilingual terminology. Various national and provincial departments and local authorities created language units to assist with language-related work such as translation, editing, interpreting and terminology harvesting.

There are at present only a few tertiary institutions in South Africa that offer official courses in terminology theory and principles. Aspects relating to these are dealt with in courses in translation or lexicography studies. Terminology training as such receives too little attention in these courses to be worthwhile. South African tertiary institutions will have to offer terminology training to people interested in this profession to solve the subject-related communication problems.

Since several language offices employ language practitioners such as translators, editors and interpreters, there is an urgent need for basic training in terminological principles and practice. They are creating new terms on a daily basis and need to acquire skills on word-forming principles, terminology management and standardisation (to name a few). At present the author of this article is providing terminology training on principles and practice of terminology and terminography to interested individuals, language bodies and language-related institutions. The multilingual dispensation requires skilled language workers, and the terminology training is aimed at enabling language workers to manage terminology in all aspects thereof.

9. The Human Language Technologies (HLT) initiative

The South African Government has approved the development of a human language technologies (HLT) virtual network. All lexicographical and terminographical endeavours will be part of the HLT virtual network. Multilingual terms will for instance be available on the HLT virtual network to end-users such as subject specialists, students, language practitioners and the general public.

10. Conclusion

The language policy of a country influences terminology development since it determines which languages to be developed. Terminology development

should adhere to the language policy and to sound terminological and terminographical principles and procedures.

Specialised communication has a central axle or hub in terminology. Terminology is therefore a strategic resource and has an important role in a country's development — especially in a multilingual country.

Effective economic, scientific and technical communication skills of the citizens of a country are developed through the use of correct and standardised terminology. The terminology practice of any country will enhance communication in various domains and will develop official (and national) languages into functional languages.

Literature

- Budin, G.** 2004. *Draft Report: National Language Services*. 15 May 2004. Unpublished report. Pretoria: NLS, DAC.
- Chin, Jeff.** 2009. <http://google-africa.blogspot.com/2009/09/google-translate-now-available-for.html>
- Cluver, A.D. de V.** 1996. *Language Development. Proceedings of a LANGTAG Workshop on Issues in Language Development: Focusing on the Future of the Apartheid Language Boards and the Harmonisation and Development of African Languages. 28 March 1996, Unisa, Pretoria*. State Language Services. DACST.
- Constitution of the Republic of South Africa.** 1996. (Act 108 of 1996). Pretoria: Government Printer.
- Department of Arts, Culture, Science and Technology.** 1996. *Proceedings of a LANGTAG Workshop on Issues in Language Development. Focusing on The Future of the Apartheid Language Boards and The Harmonisation and Development of African Languages. 28 March 1996*. Pretoria: State Language Services, DACST.
- Department of Arts, Culture, Science and Technology.** 1996. *Towards a National Language Plan for South Africa. The final LANGTAG Report*. Pretoria: DACST.
- English National Language Body, PanSALB.** 2007. *English and Multilingualism in South African Society: Papers from the English National Language Body*. Unpublished document. Pretoria: PanSALB.
- Haugen, E.** 1966. Dialect, Language, Nation. *American Anthropologist* 68: 922-935.
http://en.wikipedia.org/wiki/Languages_of_Africa
- Mayevu, G.S.** 1996. Development of the South African Literatures with Special Reference to Xitsonga. Department of Arts, Culture, Science and Technology. 1996. *Proceedings of a LANGTAG Workshop on Issues in Language Development. Focusing on The Future of the Apartheid Language Boards and The Harmonisation and Development of African Languages. 28 March 1996*. Pretoria: State Language Services, DACST.
- Mojela, M.V.** 2007. *Translators: Conveying the Message*. Paper read at the International Translation Day celebration 2007: Don't Shoot the Messenger! George Bou Conference Centre, 29–30 October 2007, Pretoria.
- Mtintsilana, P.N. and R. Morris.** 1988. Terminography in African Languages in South Africa. *South African Journal of African Languages* 8(4): 109-113.

- National Language Policy Framework (NLPF) and Implementation Plan.** 2003. Pretoria: DAC.
- Ntshangase, D.** 1996. *Working Document: Future of the Present Language Boards*. LANGTAG Subcommittee on the Development of (South) African Languages, 20 February 1996. Unpublished document. Pretoria: DAC.
- PanSALB.** 2001a. *Language Use and Language Interaction in South Africa. A National Sociolinguistic Survey*. Pretoria: PanSALB.
- PanSALB.** 2001b. *Guidelines for Language Planning and Policy Development*. Pretoria: PanSALB.
- Pan South African Language Board Act.** 1995. (Act No 59 of 1995). Pretoria: Government Printer.
- Random House.** 1987. *Random House Dictionary of the English Language*. Second Edition. New York: Random House.
- Unit for Language Management.** 2007a. *SA Language Rights Bulletin 2(1)*, September 2007. Prepared for the Pan South African Language Board by the Unit for Language Management, University of the Free State.
- Unit for Language Management.** 2007b. *SA Language Rights Bulletin 2(2)*, September 2007. Prepared for the Pan South African Language Board by the Unit for Language Management, University of the Free State.
- Webb, V.** n.d. *The Role of Language in Economic Development*. Unpublished document.
- Wright, Laurence.** 2007. Language Policy and Planning: General Constraints and Pressure. English National Language Body, PanSALB. 2007. *English and Multilingualism in South African Society: Papers from the English National Language Body: 6–11*. Unpublished document. Pretoria: PanSALB.

Terminologiebestuur in Suid-Afrika: 'n Ideaalmodel*

Dina Ferreira, *Skool vir Kommunikasiewetenskappe, Sentrale Universiteit vir Tegnologie, Bloemfontein, Suid-Afrika (dferreir@cut.ac.za)*

en

Theodorus du Plessis, *Departement van Taalbestuur en Taalpraktyk, Universiteit van die Vrystaat, Bloemfontein, Suid-Afrika (dplesslt.hum@ufs.ac.za)*

Opsomming: Die nuwe taalbestel in Suid-Afrika het tot gevolg gehad dat die ontwikkeling, die uitbouing, die bestuur en ook die verspreiding van terminologiese inligting in elf amptelike tale gedoen moet word. 'n Veranderde terminologiebestuursprogram moet derhalwe daargestel word om aan die behoeftes van 'n nuwe terminologiese gemeenskap te voldoen. So 'n model sou ideaal gesien binne die Nasionale Taaldiens se Terminologiese koördineringsafdeling geïmplementeer kon word. In die eerste gedeelte van hierdie artikel word 'n model voorgestel van hoe terminologiebestuur suksesvol binne 'n terminologiebestuursentrum aangepak kan word. In die tweede gedeelte van die artikel word die implikasies van die model in die Suid-Afrikaanse situasie verreken om vas te stel hoe die terminologiebestuursprogram aangepas sou kon word om aan die eise van die nuwe taalbestel te voldoen en om die nuwe teikengroep(e) voldoende te bereik.

Sleutelwoorde: AMPTELIKE TALE, BEPLANNING EN KOÖRDINERING, DISSEMINERING, EVALUERING, GEMARGINALISEERDE TALE, INHEEMSE TALE, IMPLEMENTERING, MEERTALIGE TERMINOLOGIEBESTUURSTEEEM, NAVORSING, ONTWIKKELENDE TALE, ONTWIKKELDE TALE, OPLEIDING, STANDAARDISERING, TERMBANK, TERMINOLOGIE, TERMINOLOGIEBESTUURSENTRUM, TERMINOLOGIEBESTUURSPROGRAM, TERMINOLOGIEONTWIKKELING, TERMINOLOGIEWERK

Abstract: Terminology Management in South Africa: An Ideal Model. The new language dispensation in South Africa has resulted in the development, the elaboration, the management and the dissemination of terminological information in eleven official languages. A modified terminology management program must be created for the needs of a new terminology community. Such a model should ideally be implemented at the Terminology Coordination Section of the National Language Service. In the first section of this article a model is proposed for successful terminology management within a terminology management centre. In the second section of the article the implication of the model is explicated within the South African situation to establish how the terminology management program should be adjusted to satisfy the needs of the new language dispensation as well as how to adequately reach the intended target group(s).

* Hierdie artikel is 'n verwerking en aanvulling van 'n gedeelte van 'n D.Phil.-proefskrif *Terminologiebestuur in Suid-Afrika met spesifieke verwysing na die posisie van historiese ingekorte tale* wat in 2002 onder die promotorskap van prof. L.T. du Plessis aan die Universiteit van die Vrystaat, Bloemfontein, Suid-Afrika, voltooi is.

Keywords: DEVELOPED LANGUAGES, DEVELOPING LANGUAGES, DISSEMINATION, EVALUATION, IMPLEMENTATION, INDIGENOUS LANGUAGES, MARGINALISED LANGUAGES, MULTILINGUAL TERMINOLOGY MANAGEMENT SYSTEM, OFFICIAL LANGUAGES, PLANNING AND COORDINATION, RESEARCH, STANDARDISATION, TERM BANK, TERMINOLOGY DEVELOPMENT, TERMINOLOGY MANAGEMENT

1. Inleiding

Die taalbehoefte wat in Suid-Afrika ontstaan het op amptelike vlak as gevolg van die politieke veranderinge sedert 1994 het tot gevolg dat oor verskillende aspekte van taalbestuur herbesin moes word. Een van die terreine waar dit veral nodig is, is terminologieontwikkeling en -bestuur. Die model vir terminologiebestuur wat in die vorige taalbestel gebruik is en tot 1998 nog toegepas is, is na aanleiding van 'n nasionale vaktaalondersoek (Cluver en Scheffer 1984) in die middel van die tagtigerjare ontwerp. Hierdie model is afgestem op 'n tweetalige taal- en terminologiebestel waar die aksent op die ontwikkeling van Afrikaans val. Ook was dit hoofsaaklik gemik op vakkundiges in spesifieke vakgebiede om sodoende Afrikaans as taal van die wetenskap en tegnologie te bevorder.

'n Meer resente ondersoek ten opsigte van die terminologiebedryf in Suid-Afrika is gedurende 2000 van stapel gestuur toe 'n nasionale terminologie-oudit gedoen is om vas te stel wat die stand van die terminologiebedryf in Suid-Afrika is (vgl. Alberts 2000b). Uit 'n beoordeling van die data wat deur hierdie oudit beskikbaar gekom het, word vasgestel dat die volgende komponente tydens die oudit ondersoek is:

- die situasie, volgens 'n situasieanalise wat daarop wys dat die Terminologiekoördineringsafdeling in die nuwe taalbestel 'n totale herstrukturering moes ondergaan tydens die amalgamering van die Nasionale Terminologiediens (NTD) en die Staatstaaldiens in die Nasionale Taaldiens en dat daar sedertdien volgens die taalgroepe by die Terminologiekoördineringsafdeling gewerk word;
- die kritieke suksesfaktore wat aantoon dat die Terminologiekoördineringsafdeling die Regering adviseer en ondersteun met die uitbreiding en bevordering van strategieë betreffende tegniese tale, dat die terminologiekoördineringsafdeling effektiewe wetenskaplike en tegniese kommunikasie bevorder deur 'n terminologiese en tegniese taaldiens te lewer, doelgerigte terminologieprodukte daarstel en die Nasionale Termbank bestuur. Daar word ook bevind dat die Terminologiekoördineringsafdeling op nasionale vlak 'n groot rol te speel het by terminologiestandaardisering en -koördinerings deur samewerking met provinsiale en ander taalkantore, dat eerstetaalterminoloë in die elf amptelike tale gebruik word om terminologiedata in 'n verskeidenheid registers te dokumenteer, dat terminologieprojekte beplan en uitgevoer word in oor-

eenstemming met gemeenskaps- en kliëntebehoefte en dat die sukses van die Terminologiekoördineringsafdeling gemeet word aan die hand van die wyse waarop tegniese en wetenskaplike woordeskat vir die voorheen gemarginaliseerde tale uitgebrei word;

- die oogmerke, wat slegs genoem word;
- die funksies en strategieë wat daarop wys dat die Terminologiekoördineringsafdeling kommunikasie en inligtingvloei bevorder deur die beskikbaarstelling van gestandaardiseerde terminologie in 'n verskeidenheid aktiwiteitsterreine in al die amptelike tale, dat die Terminologiekoördineringsafdeling 'n gratis terminologienavraagdiens lewer, riglyndokumente saamstel vir standaardisering van terminologie en terminografie, indiensopleiding aan terminoloë en terminograwe verskaf, tegniese taalinligting beskikbaar stel aan geletterdheidsprogramme, tegniese taalontwikkelingsprojekte koördineer en fasiliteer, navorsing op 'n verskeidenheid gebiede doen, terminologiedata op die MultiTerm-databasis dokumenteer en dissemineer, die Nasionale Termbank bestuur, rekenaarbystand verleen aan 'n aantal terminologieprojekte buite die Terminologiekoördineringsafdeling en in 'n fasiliteringskapasiteit optree deur ondersteuning, projekfasilitering en skakeling met buite-instansies;
- 'n lys van tweetalige woordeboekuitsette deur die Nasionale Vaktaal-diens (NVD) en NTD vanaf 1959;
- 'n lys van die lopende projekte by die Terminologiekoördineringsafdeling tydens 2000;
- 'n lys van ou NTD-projekte wat nog nie afgehandel is nie;
- 'n lys van die buiteprojekte deur privaat persone of organisasies waarby die Terminologiekoördineringsafdeling betrokke is;
- die werkwyse by die Terminologiekoördineringsafdeling, wat die dokumentering en dataprozessering (werkvloei), die gebruik van vakkomitees, standaardisering en disseminering by die Terminologiekoördineringsafdeling uiteensit; en
- skakeling op lokale, provinsiale, nasionale en internasionale vlak, wat daarop wys dat die Terminologiekoördineringsafdeling daarop fokus om 'n diens te lewer aan die vak- en taalgemeenskap van die land aangesien die kliënte van hierdie afdeling uit 'n wye verskeidenheid persone bestaan en dat die Terminologiekoördineringsafdeling lid is van verskeie oorsese terminologieorganisasies en ook een van die streekskorrespondente en kolomredigeerders van die Internasionale Federasie vir Terminologiebestuur se internasionale publikasie NEOTERM is.

Die oudit lig ook 'n aantal sterk en swak punte by die Terminologiekoördineringsafdeling uit en sekere aanbevelings word ten slotte gemaak.

Die sterk punte by die Terminologiekoördineringsafdeling is volgens Alberts (2000b) geleë in die infrastruktuur van die afdeling. Aanpassings en herstrukturering is by die Terminologiekoördineringsafdeling op die bestuursvlak en by die werkmethode gemaak om by die behoeftes van die terminologiegemeenskap in die nuwe taalbestel aan te pas. Die Terminologiekoördineringsafdeling het as deel van die aanpassings die TRADOS MultiTermprogram geïmplementeer wat gebruik maak van die gevorderdste tegnologie vir die dokumentering van terminologie in enige aantal tale.

Die swak punte word soos volg geïdentifiseer:

- daar bestaan nie dieselfde mate van ondervinding by terminologiebestuur in al die amptelike tale nie;
- daar word 'n probleem ondervind om effektiewe toegang tot vakkundiges en tekskorpora in die aktiwiteitsterreine en -domeine te verkry;
- daar bestaan nie 'n effektiewe terminologiebeleid om terminologiebestuur by die Terminologiekoördineringsafdeling te lei nie, en
- daar bestaan 'n tekort aan goed opgeleide terminoloë by die Terminologiekoördineringsafdeling.

Alberts (2000b) maak aanbevelings ten opsigte van:

- die Terminologiekoördineringsafdeling as klaringshuis vir terminologie-werk in Suid-Afrika;
- die opleiding van personeel en medewerkers by die Terminologiekoördineringsafdeling; en
- die akkreditering van terminoloë en terminografe deur die Suid-Afrikaanse Vertalersinstituut (SAVI).

Sy skenk egter nie pertinent aandag in die oudit aan die daarstel van 'n model vir Terminologiebestuur by die Terminologiekoördineringsafdeling nie.

Die oogmerk met hierdie ondersoek is juis om 'n model te ontwikkel vir terminologiebestuur in Suid-Afrika wat rekening hou met die eise van die nuwe taalbestel en om die nuwe teikengroep(e) voldoende te bereik. Dit sal gedoen word aan die hand van internasionale benaderings tot terminologiebestuur.

2. Terminologiebestuur

Galinski en Budin (1993: 209) beskryf terminologiebestuur as 'n konsep parallel aan inligtingsbestuur waar beide inligtingsbestuur en terminologiebestuur onmisbare instrumente is binne 'n omgewing waar daar gewerk word met kwaliteitbestuur.

Wright en Budin (1997: 2) verduidelik terminologiebestuur as enige doelbewuste manipulerings van terminologiese inligting. Hulle lig drie aspekte uit, te wete praktiese terminologiebestuur, die sistematiese versameling van terminologiese inligting en die *ad hoc*-onttrekking van sodanige inligting.

Ferreira (2002) se definisie sluit aan by bestaande benaderings in die literatuur rondom terminologiebestuur:

Terminologiebestuur is [noodsaaklik] om terminologiese inligting te organiseer en te koördineer as voorwaarde vir die effektiewe verspreiding van kennis en die daarstel van hoëgehalte-terminologiese inligting. Die primêre doelwit met terminologiebestuur behels derhalwe die skep van werk- en ander omstandighede om 'n hoëgehalteterminologie-uitset te kan lewer. Terminologiebestuur plaas met ander woorde 'n hoë premie op produktiwiteit en gehalte.

Sy behandel terminologiebestuur as 'n komplekse verskynsel wat uiteengesit word aan die hand van sekere bestuursfunksies wat uitgevoer moet word. Daar bestaan altyd 'n wisselwerking tussen die verskillende bestuursfunksies. Hierdie wisselwerking word in Figuur 1 skematies voorgestel.

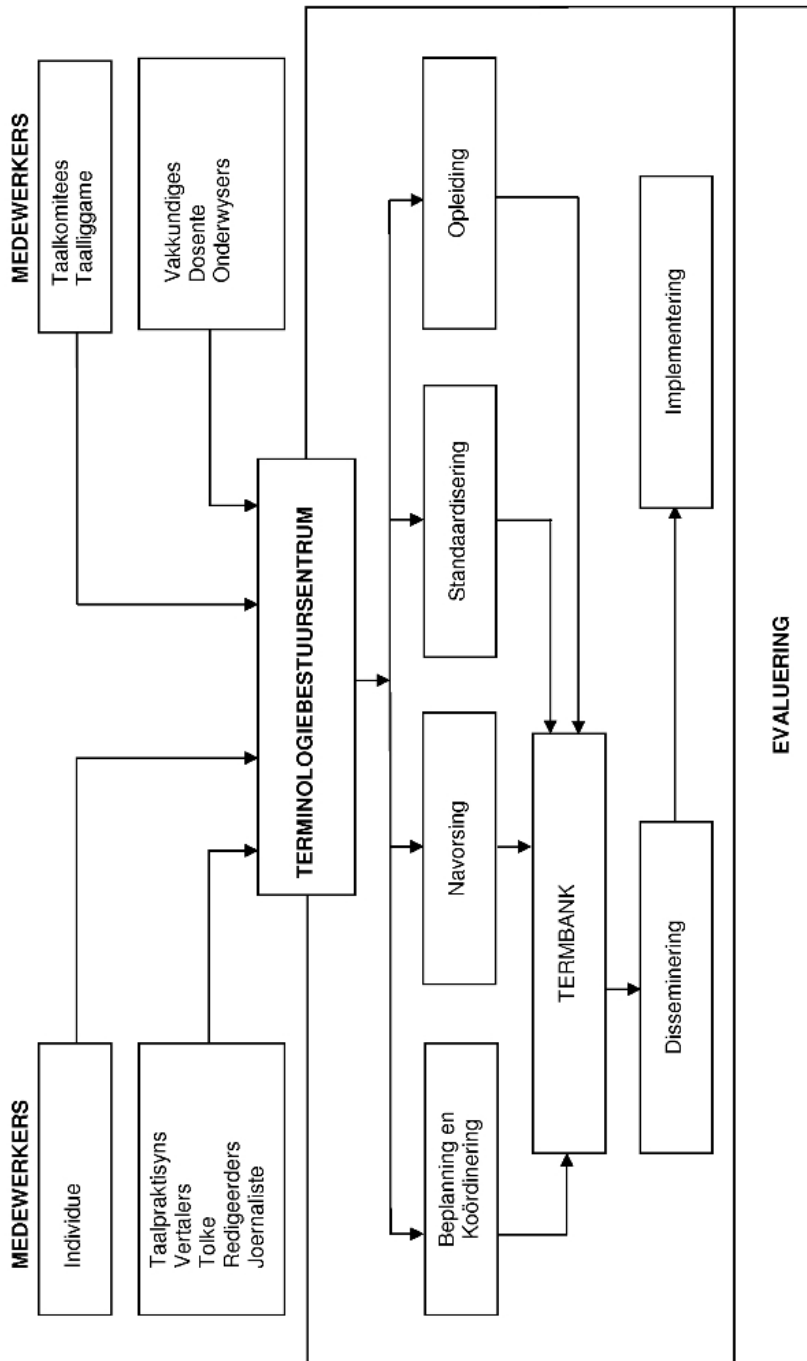
Alhoewel die belangrikheid van die bestuursfunksies sal wissel na mate van die stadium waarin terminologiebestuur in die gemeenskap verkeer, moet georganiseerde terminologiebestuur bestaan uit die verspreiding van die verantwoordelikhede en take. In gemeenskappe waar georganiseerde terminologieaktiwiteite voorkom, word terminologiebestuur oor die algemeen georganiseer na aanleiding van die gemeenskap se sosiale en politieke struktuur, die oogmerke en doelwitte wat met terminologiebestuur bereik moet word en die finansiële ondersteuning wat beskikbaar is.

Selfs al beskik 'n gemeenskap oor 'n terminologiesentrum waar al die aktiwiteite gekoördineer word, moet die terminologiesentrum ondersteun word deur medewerkers (individue, taalpraktisyne, vertalers, tolke, redigeerders, joernaliste, taalkomitees, taalliggame, vakkundiges, dosente, onderwysers). Hierdie medewerkers speel veral 'n belangrike rol by die standaardisering van nuwe terme en om prioriteitsgebiede vir terminologieontwikkeling te identifiseer.

2.1 Beplanning en koördinerings van terminologiebronne

Deurlopende en voortdurende beplanning en koördinerings is van uiterste belang by terminologiedata aangesien terminologie midde-in 'n vinnig veranderende omgewing ontstaan en aangepas moet word by die veranderende behoeftes binne die gemeenskap. Beplanning op strategiese vlak is van kardinale belang.

Die hoofdoel met die beplanning en koördinerings van terminologiedata is om werk- en ander omstandighede te skep wat die rolspelers in staat stel om konstant 'n hoëkwaliteituitset te lewer. Beplanning en koördinerings plaas met ander woorde 'n hoë premie op produktiwiteit en produktiwiteit kan verbeter



Figuur 1: 'n Skematiese voorstelling van sistematiese Terminologiebestuur in 'n Terminologiebestuursentrum

word indien dit gemeet word. Produksienorme is noodsaaklik vir die meting van produktiwiteit. Uitsette is veral belangrik gedurende die beplannings- en koördineringsproses aangesien uitsette gekontroleer en geëvalueer kan word. Die doelwitte met die versameling en disseminering van terminologie het daarom 'n baie belangrike invloed op die werkmetode wat by terminologiebestuur binne 'n spesifieke omgewing gebruik word (Ferreira en Du Plessis 2007: 56).

Ferreira en Du Plessis (2007) beoordeel drie werkmetodes waarvolgens terminologiedata tans suksesvol beplan en gekoördineer kan word en kom tot die slotsom dat uitkomst na gelang van die spesifieke werkmetode kan verander. Hierdie verandering vereis dat terminologiedoelwitte en -behoefes as deel van die strategiese plan geïdentifiseer moet word, wat weer verseker dat proaktiewe bestuur toegepas word.

Die beplanning en koördinerings van terminologiebronne kan egter nie alleen die sukses van die terminologiesentrum verseker nie. Hierdie elemente moet ondersteun word deur gesistematiseerde terminologiesnavorsing om die saamstel van sistematies gestruktureerde terminologieë te verseker.

2.2 Terminologiesnavorsing en die saamstel van sistematies gestruktureerde terminologieë

Aangesien die doelwit met terminologie-inligting nie altyd dieselfde binne verskillende gemeenskappe is nie, moet die rol van terminologiesentrum vasgestel word om te bepaal watter vorm die navorsing moet aanneem.

Die analise van terminologiesregulering in verskillende gemeenskappe wys op twee basiese benaderings: 'n preskriptiewe benadering (waarvan die doelwit is om die gebruik te beïnvloed deur spesifiek dié terme aan te dui waarvan die gebruik gemerk of gestandaardiseer is), en 'n minder rigiede benadering wat die gebruik van 'n term aanbeveel eerder as dikteer, die deskriptiewe benadering.

2.2.1 Die deskriptiewe terminologiesbenadering

Die deskriptiewe benadering behels dokumentering deur vakkundige en nie-vakkundige terminoloë, wat terminologiese produkte en aanlynbronne vir gebruik deur vakkundiges en taalpraktisyne (soos vertalers, tegniese skrywers en studente) voorsien. Die doelstelling met hierdie benadering is om die terme, wat aan konsepte binne 'n spesifieke veld toegesê is, te dokumenteer, maar dit stel nie die gebruik in die domein vas nie. Die deskriptiewe benadering se produkte is met ander woorde nie voorskriftelik van aard nie, maar al die terme wat voorkom of voorgestel word vir 'n konsep, word gedokumenteer (vgl. Wright en Budin 1997: 18).

In baie velde is die konsepte wat benodig word onpresies en vloeibaar. Daar is met ander woorde nog nie ooreengekom oor die terme wat gebruik moet word om die konsepte te benoem nie. Dit is juis binne hierdie kontekste

waar die deskriptiewe benadering benodig word om konsepte te benoem, en die plek van gebruik te organiseer.

Die uitkoms met terminologiebronne wat volgens die deskriptiewe benadering saamgestel is, veronderstel dat uitgebreide navorsing binne tale gedoen word om versamelde materiaal beskikbaar te stel wat sal dien as die terminologieskorp van die vakgebied of domein.

2.2.2 Die preskriptiewe terminologiebenadering

Hierdie benaderingsmetode ondersoek praktyke en metodologieë in die normatiewe terminologie en werk primêr binne die raamwerk van nasionale en internasionale standaardisasie, soos die Internasionale Organisasie vir Standardisering en professionele verenigings binne vakgebiede.

Die versameling van preskriptiewe terminologie-inligting word gewoonlik aangepak deur vakkundiges in standaardiseringskomitees, nomenklatuurspesialiste en taalbeplanners. Preskriptiewe terminologie-inligting behels dat daar oor betekenis ooreengekom word vir tegniese spesifikasies en standaarde en dat terminologie in die vorm van gestandaardiseerde terminologie geïmplementeer word. Die ideaal binne gestandaardiseerde terminologieswerk is geharmoniseerde terminologie wat op 'n rekenaarnetwerk in soveel moontlik tale aanlyn beskikbaar gemaak kan word, en deur almal wat dokumente saamstel of data bestuur, gebruik kan word.

Die uitkoms van preskriptiewe terminologieswerk veronderstel dat gestandaardiseerde terminologie-inligting in soveel tale moontlik aanlyn beskikbaar is.

2.3 Die standaardisering van terminologie-inligting

Sager (1990: 114) beskryf termstandaardisering as 'n proses wat daaruit bestaan dat die gebruikers van 'n vakgebied of domein konsensus bereik om 'n gegewe term vir gebruik in spesifieke omstandighede aan te wend.

Standaardisering het as doelwit die minimalisering van oortollige definisies en die verbeterde koördinerende van terminologie. Termstandaardisering sluit ook in dat norme behoorlik verstaan en presies geïnterpreteer sal word om die standaardisering van terminologie te bevorder in verslae en ander tegniese dokumente en om die betekenis van tegniese terme te verduidelik aan persone wat nie met hulle bekend is nie.

Die fundamentele streef by terminologiestandaardisering is die standaarddefinisie. Definisies word benodig vir alle ondubbelsinnige interpretasies van standaardprosedures, metodes of praktyke.

Koördinerende liggame bestaan dan ook om terminologie uit verskillende vakgebiede te standaardiseer. Voorbeelde van internasionale standaardiseringsliggame wat hierdie funksie verrig, is die Internasionale Elektrotegniese Kommissie (IEC), tans die alleenliggaam met verwysing na standaarde in elek-

triese ingenieurswese en elektronika, en die Internasionale Organisasie vir Standaardisering (ISO).

Die Tegniese Komitee 37, Terminologiebeginsels en Koördinerings van die Internasionale Organisasie vir Standaardisering (ISO/TC37) voorsien standaarde vir terminologiedata aan internasionale standaardiseerders. ISO/TC37 het in 1992 'n dokument gepubliseer wat bekend staan as die *International Terminology Standards — Preparation and Layout* (ISO 1024: 1992). Hierdie dokument stel riglyne vas vir gebruik by die voorbereiding en uitleg van internasionale standaarde.

Dit is egter nie voldoende om terminologiese beginsels uit te brei op internasionale vlak, of om hulle te publiseer as ISO-standaarde, en om hulle by te werk deur hersiening nie. Alle pogings tot standaardisering van terminologiese werk is doelloos indien die gestandaardiseerde terme nie geïmplementeer, versprei en gebruik word nie.

Die uitkoms wat deur standaardisering veronderstel word, berus daarop dat terminologiestandaardisering as 'n soort riglyn dien vir die meerderheid van die persone wat geïnteresseerd is in die verenigde regulering van 'n spesifieke vakgebied of domein. Deur terminologiestandaardisering word 'n sekere uniformiteit gewaarborg en optimale professionele kommunikasie word verseker.

2.4 Die opleiding van terminoloë, terminograwe en medewerkers

Volgens Cabré (1999: 220) kan die opleiding van terminoloë nie losgemaak word van die sosiale konteks waarin die terminoloë werk nie. Dit beteken dat die terminologiese wetenskap in diens staan van die taalbeleid van 'n gemeenskap en noodsaaklik is vir die uitvoering daarvan. Dit het ook tot gevolg dat die opleiding van professionele terminoloë georganiseer moet word in samewerking met die spesifieke terminologiese behoeftes van die gemeenskap waarvoor die terminologie saamgestel word.

Terminologiese opleiding wat tans baie pragmaties en toepassingsgeoriënteerd blyk te wees, kan derhalwe as 'n diensvaardigheid beskryf word wat ten beste deur praktiseerders met direkte ondervinding oorgedra kan word.

Die inhoud van kursusse vir terminologiese opleiding moet daarom gebaseer word op 'n behoefteanalise in die gemeenskap waarvoor die terminologie saamgestel word, en die praktyk moet 'n rol vertolk by die beplanning en vasstelling van die kursusinhoud van terminologiese opleiding.

'n Ander belangrike aspek wat by opleiding in ag geneem moet word, is die opleiding van medewerkers. Cabré (1999: 220) wys daarop dat die opleiding van praktisyns in terminologiedata, nie verwar moet word met die opleiding van terminoloë nie. Behalwe die opleiding van terminoloë is daar ook 'n behoefte aan die opleiding van vertalers, tolke, tegniese skrywers, wetenskaplikes, tegnisi, inligting- en dokumentasiespesialiste, taalbeplanners, en rekenaarwetenskaplikes wat spesialiseer in kunsmatige intelligensie van toepassing

in tale. Hierdie persone benodig opleiding in terminologiedata om spesifieke terminologiese probleme die hoof te bied wat ontstaan wanneer daar met meer as een taal gewerk word, 'n spesiale woordeskat gebruik word of werk aan meertalige terminologienavorsing onderneem word.

Opleiding dien derhalwe ook om die leemtes in die terminologiebestuursproses te identifiseer waarna die elemente waarby daar probleme bestaan, herbeplan en herimplementeer moet word.

2.5 Rekenaargesteunde terminologiebestuur en termbanke

Deesdae word dit algemeen aanvaar dat die enigste praktiese manier waarop leksikale data bestuur kan word, met behulp van die rekenaar is. Die waarde van die gebruik daarvan wat betref spoed, buigsaamheid en bergingskapasiteit het daartoe gelei dat daar 'n groeiende neiging is na die outomatisering van terminologiese dataprozessering. Nuwe konsepte ontstaan met nuwe ontdekkings en uitvindings as resultaat van navorsing en ontwikkeling. Koherente en sistematiese metodes by die formulering van terme word daarom al hoe belangriker vir die daarstel van betroubare terminologiestandaarde. Standaardterminologieë moet voorberei word op die basis van terminologiebeginsels en -metodes, wat geneem word uit die bevindinge van die terminologiewetenskap aan die een kant en praktiese ondervinding van terminologiewerk aan die ander kant. Terminologiebanke kan nie doeltreffend bestuur word sonder samewerking deur die verskillende rolspelers nie. Dit veronderstel ook die toepassing van internasionale standaarde.

Tans is daar 'n wye reeks programmatuur en ander instrumente beskikbaar om die strukturering, berging, uitruiling en disseminering van terminologiedata te behartig. Die voordele van hierdie sisteme is vanselfsprekend:

- vinnige aanpassing by gebruikersbehoefte en die vinnige veranderende werklikheid;
- die moontlikheid om nuwe terminologiedata te versamel en te dissemineer soos wat dit geskep word, om duplisering, verdraaiing en misverstande uit te skakel;
- die moontlikheid om terminologiedata wat benodig word vir die skryf van produk- of diensdokumentasie, in een of meer verskillende taalmoontlikhede daar te stel; en
- die moontlikheid om gestandaardiseerde terminologiedata te gebruik om terminologiekonstantheid in dokumente en hul vertalings te waarborg.

Terminologieskeppingsagente beskik tans oor 'n groot arsenaal van instrumente om hulle in hul werk by te staan. Ontwikkelde standaardtale wat reeds 'n gerekenariseerde databasis gebruik, het oor die algemeen 'n groot versameling data in masjienleesbare vorm beskikbaar wat dit moontlik maak om inlig-

ting te verskaf oor die terminologieskeppingspatrone van enige dissipline. Ontwikkelende tale probeer ook toenemend om terminoloë wat vaardig is in rekenaartegniese, op te lei vir die versameling en disseminering van terminologiedata.

2.6 Disseminering van terminologiedata

In die era van kommunikasie het dit duidelik geword dat terminologiebestuur met behulp van die rekenaar groot voordele inhou, veral ten opsigte van verhoogde produksie en vakkundige gehalte. Die totstandkoming van termbanke vloei hieruit voort. 'n Termbank verleen aan die gebruiker vinnige, goedkoop en gerieflike toegang tot terminologiedata. Indien die termbank deel van 'n inligtingsnetwerk vorm, kan die gebruiker oor nasionale en selfs internasionale grense heen toegang tot ander databasisse en ander netwerke verkry.

Die tendens regoor die wêreld ten opsigte van die disseminering van terminologiedata is spesifiek daarop gemik om terminologiebronne nader aan mekaar te bring. Kleiner en individuele gebruikers deel en ruil hul gespesialiseerde terminologieversamelings uit en groter sentra dissemineer en maak hul terminologiedata beskikbaar deur middel van termbanke. Aangesien inligting oor wetenskaplike en tegnologiese innobering op groter skaal as ooit tevore aangebied word, veral deur die gebruik van moderne media, is daar 'n toeneemende behoefte aan die disseminering van terminologiedata aan die algemene publiek. Met die uitbreiding van kennis deur opleiding in verskillende domeine is daar voortdurend meer individue in die gemeenskap wat beter opgelei is, en daar bestaan 'n al groter wordende behoefte om kennis met betrekking tot terminologiedata aan 'n wyer publiek te verskaf. Veral in die industrie is daar 'n groeiende behoefte om toegang te verkry tot terminologiedata van 'n hoë standaard en benewens die daarstel van die nodige infrastruktuur om die gebruikers en lewensinsigters van terminologie-inligting nader aan mekaar te bring, is dit ook die taak van die terminologiebestuursentrum om die versamelde terminologie-inligting aan alle potensiële gebruikers daarvan beskikbaar te stel.

2.7 Implementering van terminologiedata

Die doelwit met die daarstel van terme is dat hulle dadelik deur die gebruikers in daardie vakgebied of domein gebruik kan word. Om dit te bewerkstellig, moet die terminologie so gou moontlik aan die gebruikers beskikbaar gemaak word.

Sommige taalakademies en taalkomitees beskik oor hul eie vaktydskrifte waarin termlyste gepubliseer word. Van hierdie vaktydskrifte is egter duur en bereik nie altyd die doelgroep waarvoor hulle bedoel is nie. 'n Goedkoper manier is om termlyste beskikbaar te stel en hulle dan aan die doelgroep te stuur vir gebruik. Om die gebruik van terme te ondersteun, moet geskrewe

dokumentasie en tekste beskikbaar gestel word. Die sirkulasie van die terme en die bevordering van die gebruik daarvan is seker die moeilikste onderafdeling van terminologiebeplanning. Elemente wat hierdie aspek ondersteun, is opvolgleeswerk en opleidingsmateriaal wat voorberei moet word sodat die terme in die konteks van gebruik voorkom. Terme moet ook deurlopend by opleiding en bespreking binne die vakgebied of domein gebruik word om penetrasie, stabilisasie en standaardisering te bewerkstellig.

In baie lande bestaan daar 'n georganiseerde terminologiese navorsingsprogram en administratiewe agentskappe of sentra vir die werk aan spesiale terme. 'n Georganiseerde terminologiese navorsingsprogram het tot gevolg dat terminologiedata geïnstusionaliseer word. Instusionalisering het die voordeel dat beter koördinerende van die werk en versnelde rasionalisasie van die ekonomiese en menslike hulpbronne moontlik gemaak word. Indien regeringsagentskappe aan die stuur van terminologiese aktiwiteite staan, kan dit tot gevolg hê dat terme 'n geregverdigde status verkry wat nie altyd moontlik is wanneer slegs nieregeringsliggame betrokke is nie. In die meeste lande met 'n aktiewe taalbeleid is die agentskappe wat verantwoordelik is vir die implementering van die taalbeleid, ook verantwoordelik vir die bestuur van terminologiedata.

Afgesien van die voordele wat regeringsbetrokkenheid by terminologiebestuursaktiwiteite inhou, kan die betrokkenheid van akademiese instellings, semi-amptelike sentrums en die privaatsektor nie geïgnoreer word nie, aangesien vakkundigheid in spesiale vakgebiede uit hierdie sektore kom. Die deelname van dié gebruikers aan die terminologieontwikkelingsproses het tot voordeel dat standaardisering nie geforseerd is nie, maar wel natuurlik plaasvind. Die gebruikers van terminologie moet voel dat hulle deel van die proses vorm om die kwaliteit van die werk en die effektiewe gebruik van die terme te verseker. Eindgebruikers moet lede van standaardiseringskomitees wees, navorsingspanne lei en aan terminologiebestuur deelneem.

2.8 Evaluering van terminologiedata

Die behoefte aan evalueringsprosedures (soos deur Rubin (1971: 221) vir taalbeplanning voorgestel) is ook van belang by terminologiebestuur. Haugen (1983) sluit evaluering onder implementering by sy taalbeplanningstipologie in.

Evaluering is egter een van die aspekte by terminologiebestuur wat in die literatuur afgeskeep word. Strategieë moet nie slegs daargestel word vir beplanning, koördinerende, navorsing, disseminering en implementering van terminologiebestuur nie; dit is net so belangrik om die kwaliteit en die sukses van hierdie strategieë te evalueer. Evaluering is teweens die meting van die kwaliteit en sukses van terminologiebestuur en is 'n deurlopende proses wat gestruktureer moet word om gereelde terugvoer te bewerkstellig sodat implementering aangepas kan word uit inligting wat vanuit die evaluering afkomstig is.

Vrae wat tydens evaluering gevra kan word, kan soos volg geformuleer word:

- Hoe effektief is die metodologie wat by terminologiebestuur gevolg word?
- Hoe bruikbaar is die materiaal wat beskikbaar gestel word?
- Is die tipe terminologiebestuur wat gevolg word, effektief?
- Word die terminologiebehoefte wat aanvanklik gestel is, bereik?

Tans bestaan daar nie vasgestelde kriteria vir die evaluering van kwaliteit by terminologiebestuur nie. Budin (1999: 7) wys op 'n aantal verskillende vlakke waarop terminologiebestuur moontlik geëvalueer kan word en stel voor dat die volgende kriteria in gedagte gehou kan word:

- Die inhoudsvlak wat te make het met die aanvaarbaarheid, die korrektheid (insluitend semiotiese toewysings), die geldigheid (konsepte, definisies, voorstelle), die toepaslikheid (situasie- en konteksafhanklikheid), die geskiktheid (in verhouding met die doelwitte gestel en gebruikersbehoefte gespesifiseer) van terminologie binne 'n spesifieke vakgebied of domein.
- Die formele vlak wat verantwoord word deur die linguistiese korrektheid, die ortografie, die morfo- en tekssintaksis, kollokasies, die grammatika, die geslag en die tipe woordsoort van die terminologie onder bespreking.
- Die koste-effektiwiteitsanalise moet wat terminologiebestuur betref, bewys lewer dat die voorgestelde plan en die implementering daarvan koste-effektief is.
- Die bruikbaarheidsvlak veronderstel dat daar bruikbaarheidtoetsing-prosedures sal wees ten opsigte van die gebruikersspesifikasies sowel as aanvaardingstoetsing vir bruikbaarheid deur gebruikers, funksionaliteit en relevansie in situasies en gebruikskontekste en die inskakeling (van gereedskap en inligtingsbronne) by die verskillende prosesse van terminologiebestuur.

Die sukses van terminologiebestuur word bepaal deur die ontwerp van evalueringstrategieë en die belangrikheid wat aan die bevindinge hiervan geheg word. Evaluering moet derhalwe op so 'n wyse aangepak word dat die resultate teruggeploeg kan word in die sisteem om tot voordeel van die gebruikers te dien.

3. Terminologiebestuur in Suid-Afrika na 1994

Politieke en sosiale verandering in Suid-Afrika het sedert 1994 die funksionele

rol van die tale in Suid-Afrika verander. Die Grondwet (Wet 108 van 1996) stipuleer dat daar elf amptelike tale in Suid-Afrika is, en dat die histories gemarginaliseerde tale ontwikkel en bevorder moet word om die funksies te kan verrig wat die moderne samelewing van hierdie tale verwag. Die Grondwet plaas veral klem op die ontwikkeling van die sogenaamde histories gemarginaliseerde tale (Artikel 6(2)) om in pas te kom met die tegnologie en om terselfdertyd die verwaarlosing as gevolg van die ideologie van die vorige taalbestel reg te stel. Die taalopdrag veronderstel:

- die skep van 'n infrastruktuur deur die staat vir die ontwikkeling en bestuur van al die tale wat in die Grondwet genoem word (soos voorsien word in die betrokke artikel), en
- die instelling van die Pan-Suid-Afrikaanse Taalraad (PanSAT) (Artikel 6(5)).

Die taalopdrag aan die staat behels dat effektiewe en toeganklike dienste verskaf moet word om die taalregte van die gemeenskap te erken en om die funksionele gebruik van die amptelike tale gelykwaardig te bestuur. Die staat het die verantwoordelikheid gekry om stappe te doen om die gebruik van die inheemse tale in hoër funksies in die samelewing te bevorder. Dit beteken dat maatreëls getref moes word om te verseker dat hierdie tale direk baat vind by programme deur die staat (wat gerig word op die bevordering van agtergeblewe gemeenskappe). Die gebruikers van hierdie tale moet bewus gemaak word van die waarde van die tale as funksionele kommunikasiemiddels en in staat wees om hierdie tale op alle vlakke van die samelewing en in alle domeine te kan gebruik as tale van kommunikasie. Wat hier vereis word, is die instel van wetgewing, ondersteun deur infrastruktuur om taalbeplanningsagente die geleentheid te bied om effektiewe terminologiebestuur te bewerkstellig.

3.1 Beleidsontwikkeling en terminologiebestuur

Van die belangrikste taalbeleidsdokumente wat die beleid en praktyk rondom terminologiebestuur na 1994 ondersteun, word saaklik genoem:

Die LANGTAG-verslag (DACST 1996) bevat 'n "Opsomming van die belangrikste aanbevelings van LANGTAG" (vgl. DACST 1996: 201-204) wat op die uitbreiding van terminologie en die bestuur van terminologie in Suid-Afrika betrekking het. Implikasies wat betref die korttermyn- en langtermynaanbevelings by terminologiebestuur het tot gevolg gehad dat daar na 'n nuwe terminologiebestuursprogram vir Suid-Afrika gekyk moes word. Daar moet ook omgesien word na die terminologiebehoefte van elf amptelike tale, met ander woorde terminologiebestuur binne 'n meertaligheidsbestel. Die terminologiegebruikersgemeenskap verskil ook geheel en al van die terminologiegebruikers in die vorige taalbestel aangesien terminologiedata op verskillende vlakke en in verskillende registers beskikbaar moet wees.

Die Staatstaaldiens en die Nasionale Terminologiediens het gedurende April 1998 geamalgameer en die Nasionale Taaldiens (NTD) binne die Departement Kuns, Kultuur, Wetenskap en Tegnologie het tot stand gekom. Hierdie kantoor moes 'n transformasie ondergaan ten einde taalfasiliteringswerksaamhede in die Staatsdiens te kan moniteer en om doelmatige dienste ten opsigte van al elf amptelike tale te kan lewer.

Die LANGTAG-verslag (DACST 1996) is deur die Nasionale Taaldiens verder gevoer en die *National Language Policy Framework* (NLPF) (DAC 2002) verskyn gedurende 2002. Terminologieontwikkeling en -bestuur word in afdelings 3.3, 3.6 en 4.9 genoem, maar nie verder gevoer nie. Die *Implementation Plan: Framework for the National Language Policy* (vgl. DAC 2003) wat 'n algemene strategie vir die ontwikkeling van die Afrikatale aanbied, word in 2003 bekend gestel, maar wat terminologiebestuur betref, word daar nie 'n spesifieke strategie uiteengesit vir die bestuur van terminologiedata binne die implementeringsplan nie.

Beide die LANGTAG-verslag en die NLPF-dokument rig egter terminologiebestuur binne die NTD in.

3.2 Organisasies en terminologiebestuur

Die NTD se Terminologiekoördineringsafdeling het sedert 1994 verskeie veranderinge ondergaan om aan te pas by die nuwe uitdaging om terminologieuitsette in al die amptelike tale beskikbaar te maak. Terwyl daar gedurende die vorige taalbestel hoofsaaklik gefokus is op terminologiedata by 'n formele register wat slegs 'n geslote gebruikersgroep ingesluit het, moet terminologiedata wat in die nuwe taalbestel benodig word, meer prakties en meer basies van aard wees om aan die behoeftes van 'n breër terminologiegebruikersgemeenskap te voldoen. Hierdie nuwe terminologiebehoefte strek oor alle domeine van kennis en menslike aktiwiteite. Meertalige terminologielyste wat sedert 1994 tot 1998 deur die Nasionale Terminologiediens (later die Terminologiekoördineringsafdeling) gepubliseer is, sluit in *Inheemse soogdiere* (1996), *Basiese gesondheidsterme* (1997) en *MIV/Vigs-bewustheidsveldtog-terme* (1998). Die tipiese termlyste wat tans saamgestel word, bevat Engelse terme as die bronsterme. Al die bronsterme word voorsien van eenvoudige verduidelikings of definisies in maklik verstaanbare Engels met termekwivalente in Afrikaans, Zulu, Xhosa, Swati, Ndebele, Tswana, Noord-Sotho, Suid-Sotho, Venda en Tsonga of 'n kombinasie van hierdie tale as doeltale (Alberts 2000b: 9-10).

Sedert Maart 2000 word daar na aanleiding van 'n versoek deur die Minister van Onderwys aan die destydse Minister van Kuns, Kultuur en Tegnologie, dr. Ben Ngubane, by die Terminologiekoördineringsafdeling gekonsentreer op die daarstel van basiese terminologiedata in geïdentifiseerde vakgebiede by onderwys (vgl. DAC 2000/2001: 79). Hierdie terminologielyste word saamgestel om leerders te help wanneer hulle in graad 5 vanaf moedertaalonderrig na Engels oorskakel. Die fokus is gerig op die uitbreiding en bevoor-

ring van terminologiedata in die amptelike tale om as ondersteunende hulpmiddels by opleiding te dien.

Die aktiwiteitsterreine wat as prioriteitsgebiede geïdentifiseer word, is Wiskunde (Mathematics), Natuurwetenskap (Natural Science), Kuns en Kultuur (Arts and Culture), Menslike, Sosiale, Ekonomiese en Bestuurswetenskappe (Human, Social, Economic and Management Sciences) en Lewensoriëntering (Life Orientation). Die projekte vir Wiskunde en Natuurwetenskap is reeds afgehandel en aan die Nasionale Taalliggame (NTL'e) vir goedkeuring en verifiëring voorgelê. Die projekte vir die Sosiale en Ekonomiese Wetenskappe wat goed op dreef is, word ten opsigte van strukturele bystand ondersteun deur PanSAT se NTL'e (*PanSALB Annual Report 2004/2005, 2006/2007*).

Buiten die onderwysprojekte word daar ook gewerk aan projekte vir Inligtingskommunikasietegnologie (in samewerking met die Departement van Kommunikasie), MIV/VIGS (in samewerking met die Departement van Gesondheid) asook 'n Parlementêre/Politieke Termlyns (*PanSALB Annual Report 2006/2007*).

Die tweede nasionale instansie wat vir terminologie in Suid-Afrika verantwoordelik is, is die Pan-Suid-Afrikaanse Taalraad (PanSAT). Hierdie instansie word grondwetlik vermeld as die liggaam primêr aanspreeklik vir taalontwikkeling.

PanSAT funksioneer as 'n onafhanklike statutêre liggaam wat gestig is ingevolge van die Pan-Suid-Afrikaanse Taalraad Wet (Wet nr. 59 van 1995) om veeltaligheid te bevorder en spesifiek om die voorheen gemarginaliseerde tale te ontwikkel. Volgens die PanSAT Wet van 1995 (soos geamendeer deur Wet nr. 10 van 1999) het PanSAT drie tipes strukture geskep vir al die amptelike tale, die Khoi- en Santale, Suid-Afrikaanse gebaretaal asook die erfenistale, te wete Provinsiale Taalkomitees, Nasionale Taalliggame en Nasionale Leksikografiese Eenhede:

- Die rol van die Provinsiale Taalkomitees (PTK's) in elke provinsie behels die bevordering en ontwikkeling van die verskillende amptelike tale binne die provinsie. Die PTK's adviseer die Raad oor enige taalaangeleenthede.
- Die Nasionale Taalliggame (NTL'e) adviseer die Raad oor enige spesifieke taal, maar in besonder oor die taalontwikkeling vir daardie taal, wat taalstandaardisering insluit, maar spesifiek ook terminologieontwikkeling vir die amptelike tale.
- Die rol van die Nasionale Leksikografiese Eenhede (NLE's) behels die samestelling van woordeboeke vir al die tale. Hul funksie by taalontwikkeling is veral leksikale modernisering. PanSAT beskik egter nie oor 'n terminologiesentrum nie en maak ten opsigte van terminologieaangeleenthede staat op advies wat deur die NTL'e voorsien word, maar ook deur opdragnavorsing daarvoor (*PanSALB Annual Report 2004/2005*).

Wat die ontwikkeling van terminologie betref, werk PanSAT en sy NTL'e saam met die Suid-Afrikaanse Standaardiseringskomitee van ISO/TC 37 (SABS TC 37) by die Suid-Afrikaanse Buro vir Standaarde (SABS). Hiervoor is Tegniese Komitees vir Terminologieontwikkeling tot stand gebring. PanSAT werk ook saam met die Departement Kuns en Kultuur (DKK) en die Departement van Onderwys om terminologie te ontwikkel in leerareas wat betrekking het op Wetenskap en Tegnologie en om strukturele bystand te verleen (*PanSALB Annual Report 2004/2005*). Dr. M. Alberts, die PanSAT-bestuurder vir Standaardisasie en Terminologie, het 'n riglyndokument opgestel vir die verifiëring en goedkeuring van terminologie en dit aan die NTL'e voorgelê. Sy is ook deurlopend besig met terminologieopleiding by verskeie instansies. PanSAT is verder ook betrokke by spesifieke navorsing ten opsigte van terminologieontwikkeling (*PanSALB Annual Report 2006/2007*).

Die NTL'e se Tegniese Komitees vir Terminologiewerk verifieer terme en verleen goedkeuring aan wat onder meer deur die NTD en ander instansies of private inisiatiewe gedoen word. Deel daarvan is om na terminologielyste te kyk en woorde wat maklik verkeerd gespél word of nuutskeppings is, te verifieer. Die Nasionale Taalliggaam vir Suid-Afrikaanse Gebaretaal is ook besig met terminologieontwikkeling. Dit maak video-opnames van nuwe terme om dié terme te probeer standaardiseer (*PanSALB Annual Report 2004/2005, 2006/2007*).

Uit die bogenoemde oorsig is dit duidelik dat beplanning en koördinerings tussen PanSAT en die NTD met betrekking tot terminologieontwikkeling en -bestuur uiters noodsaaklik is aangesien duplisering van take rondom hierdie kwessie maklik kan voorkom.

4. Terminologiebestuur in Suid-Afrika — implikasies van 'n nuwe model

Die toepasbaarheid van die model by terminologiebestuur wat in afdeling 2 voorgestel word, bied die moontlikheid om betekenisvolle insae te verkry in die ontwikkeling van 'n sistematiese terminologiebestuursprogram vir 'n meertalige Suid-Afrika. Die implikasies van die model sal vervolgens verreken word deur van die kriteria in Figuur 1 gebruik te maak.

Terwyl daar gedurende die vorige taalbestel hoofsaaklik gefokus is op terminologiedata by 'n formele register wat slegs 'n geslote gebruikersgroep ingesluit het, is terminologiedata wat in die nuwe taalbestel benodig word, meer basies van aard om aan die behoeftes van 'n breër terminologiegebruikersgemeenskap te voldoen. Daar word hoofsaaklik gekonsentreer op die daarstel van terminologiedata in geïdentifiseerde vakgebiede by onderwys sowel as die daarstel van basiese terminologiedata vir MIV/VIGS, Inligtingskommunikasietegnologie en die Parlement (DAC 2006/2007). Terminologielyste word na samestelling na die NTL'e se Tegniese Komitees vir Terminologieontwikkeling gestuur waar verifiëring en goedkeuring van die terme plaasvind.

Deurlopende beplanning en koördinerings van terminologiebronne op strategiese vlak blyk van kardinale belang te wees in die nuwe taalbestel om aan te pas by die veranderende behoeftes van die nuwe terminologiegemeenskap. Voordat beplanning op strategiese vlak egter kan plaasvind, moet daar eers vasgestel word wat die werkmethode is waarvolgens terminologie in Suid-Afrika in die nuwe taalbestel bestuur word. Ferreira en Du Plessis (2007) ondersoek drie heersende terminologiese werkmodes by terminologiebeplanning, terminologiekoördinerings en terminologiebestuur en kom tot die slotsom dat die taalbeplanningsgeoriënteerde werkmethode die aangewese metode vir die huidige Suid-Afrikaanse situasie is. Hulle bevind verder dat die suksesvolle toepassing van die taalbeplanningsgeoriënteerde werkmethode tot gevolg kan hê dat terme versamel, gesistematiseer en voorberei word vir gebruik op verskillende vlakke en registers en dat gebruikers dan in staat sal wees om makliker en doelmatiger in die amptelike tale te kan kommunikeer. Terminologie-doelwitte en -behoefte moet met ander woorde bepaal en as deel van die strategiese plan geïdentifiseer word sodat proaktiewe bestuur toegepas kan word.

Terminologienavorsing het ook in die nuwe taalbestel 'n verandering ondergaan aangesien terminologie-inligting in al die amptelike tale benodig word om die gebruikers van hierdie tale in staat te stel om die tale as funksionerende tale van kommunikasie te gebruik. Meertalige vertalende en definieringswoordeboeke en -woordelyste is tans in aanvraag om aan die behoeftes van die gebruikers van die voorheen gemarginaliseerde tale te voldoen met die oog op veral onderrig en opleiding (DAC 2006/2007). Terminoloë spesialiseer ook nie meer soos in die verlede in spesifieke vakgebiede nie, maar wel in die spesifieke amptelike tale (DACST 1999: 130). Engels word reeds in die meeste vakgebiede en domeine as kommunikasiemiddel (brontaal) gebruik. Konsepte van 'n aktiwiteitsterrein of domein word versamel deur brontaalterme te dokumenteer, konsepte te definieer en, nadat ooreengekom is oor die geldigheid van die konsep-termverhouding in die brontaal, word termekwivalente in die ander amptelike doeltale verskaf (Alberts 2000b: 22-24). Die deskriptiewe benadering by terminologienavorsing blyk hier veral van waarde te wees aangesien terme wat gebruikers verkies, gedokumenteer of versamel word om inligting beskikbaar te maak oor konsepte en om die plek van gebruik vas te stel en te organiseer, maar nie noodwendig voorskriftelik op te tree nie.

Gestandaardiseerde terminologiedata in die nuwe taalbestel word benodig om al die amptelike tale in staat te stel om die status van funksionerende tale van kommunikasie te verkry en so ontwikkeling op ander gebiede moontlik te maak en te ondersteun. Eentalige of meertalige terminologielyste wat deur die Terminologiekoördineringsafdeling of enige ander liggaam saamgestel word, word aan die Tegnieke Komitees vir Terminologieontwikkeling van die Nasionale Taalliggame (NTL'e) vir verifiëring en goedkeuring voorgelê. Die NTL'e word as die hoogste gesag in hul onderskeie tale geag en enige inrigting of individu mag terminologie direk aan hulle voorlê vir verifiëring en goedkeuring (DAC 2006/2007). PanSAT en sy NTL'e werk ook saam met die

Suid-Afrikaanse Standaardiseringskomitee van ISO/TC37 (SABS TC 37) by die Suid-Afrikaanse Buro vir Standaarde (SABS) (*PanSALB Annual Report 2004/2005*).

Wat opleiding in terminologiewerk betref, word twee verskillende tipes geïdentifiseer: die opleiding van terminoloë en terminograwe, en die opleiding van medewerkers (taalpraktisyns of vakspecialiste). Aangesien opleiding in terminologiewerk tans baie pragmaties en toepassingsgeoriënteerd is, moet praktiseerders met ondervinding vir die opleiding gebruik word. Kursusse vir terminologieopleiding moet daarom ook gebaseer word op 'n behoefteanalise in die gemeenskap waarvoor die terminologie saamgestel word om sodoende doelmatig te wees. Uit onderhoude wat gedurende Januarie 2001 met die hoof-taalpraktisyns en die terminoloë en terminograwe van die verskillende amptelike tale by die Terminologiekooördineringsafdeling gevoer is, is vasgestel dat indiensopleiding hoofsaaklik op 'n een-tot-een-basis, maar soms ook in groepverband aangebied word (Alberts en kollegas 2001). Alhoewel die handleiding wat vir interne opleiding gebruik word, nie ter insae was nie vanweë die verbod op die beskikbaarstelling van interne dokumente van die Nasionale Taaldiens, is daar uit die onderhoude vasgestel dat die handleiding wat gedurende die vorige taalbestel saamgestel is, bygewerk en aangepas is om aan die behoeftes van die terminoloë en terminograwe van al die amptelike tale te voldoen. Uit die onderhoude kon ook vasgestel word dat die opleiding wat terminoloë en terminograwe by die Terminologiekooördineringsafdeling kry, in 'n groot mate ooreenstem met die opleiding wat medewerkers van terminologiewerk gedurende 2001 ondergaan het. Die opleiding van terminoloë en terminograwe is egter baie intensiewer, strek oor 'n langer tydperk en dek volgens die mondelinge bronne alle aspekte van terminologiewerk (vgl. ook Alberts 2000a: 247). Die terminologie-oudit van 2000 (Alberts 2000b) wys daarop dat die opleidingskomponent van terminoloë en terminograwe uitgebrei kan word deur veral internasionale opleidingsbronne te gebruik.

Die tweede tipe opleiding wat aangebied word, het in die nuwe taalbestel 'n uiters belangrike funksie. Die oogmerk met terminologieopleiding vir medewerkers en rolspelers by terminologiewerk sluit in om koördinering by die daarstel van terminologieë te bewerkstellig, om samewerkingsooreenkomste met medewerkers en rolspelers te sluit en om uiteindelik gestandaardiseerde termlyste aan die terminologiegebruikers te dissemineer wat aan hulle terminologiebehoefte voldoen (DAC 2001). Daar is gedurende 2001 met die opleiding van medewerkers en rolspelers begin en terminologieopleiding is by sentrale sentra in die verskillende provinsies regoor Suid-Afrika aangebied (DAC 2001). Lede van die tegniese komitees vir terminologieontwikkeling van die verskillende NTL'e word deurlopend opgelei. Hulle word toegerus met die nodige agtergrondkennis oor die beginsels en praktyk van terminologiewerk sodat die NTL'e die verifiëring en goedkeuring van terminologie in die verskillende tale kan waarneem. Dr. M. Alberts van PanSAT bied opleiding aan vir lede van hierdie komitees oor 'n verskeidenheid aspekte van die teorie van terminologie, beginsels en praktyk van terminologiewerk en woordvormings-

beginsels. Opleiding in die beginsels en praktyk van terminologiewerk word ook by verskillende instansies deur Dr. Alberts aangebied. 'n 'TermTrain — training the trainer' werkswinkel georganiseer deur TermNet (Oostenryk) in samewerking met SABS TC 37, PanSAT, die Nasionale Taaldiens, die Departement van Kuns en Kultuur en AFRILEX is gedurende Maart 2006 aangebied (*PanSALB Annual Report 2006/2007*).

Alhoewel die uitkoms ten opsigte van praktykgerigte opleiding en gereelde terugvoer deur praktiseerders miskien nog nie ten volle bereik word nie, is die pogings om nuwe bande te smee, samewerkingsooreenkomste te sluit en medewerkers en rolspelers op te lei 'n stap in die regte rigting. Hierdie opleidingspogings kan gesien word as die eerste stap om te probeer om terminologieopleiding in voeling met die praktyk te hou.

'n Meertalige terminologiebestuursisteem vereis dat basiese terminologiedata in elf amptelike tale, wat in die behoeftes van 'n wye verskeidenheid terminologiegebruikers moet voorsien, beskikbaar gestel moes word. Om in hierdie wye verskeidenheid van terminologiebehoefte te voorsien, het die Nasionale Termbank gedurende 1998 die Trados MultiTerm-program begin gebruik (DACST 1998: 27). Die MultiTerm-program is in staat om uitgebreide terminologiese inligting oor terme te dokumenteer en dit dan elektronies te dissemineer. Alle aktiwiteite met betrekking tot die ontwikkeling van rekenaarsisteme vind plaas in oorleg met standarde soos vasgestel deur die Internasionale Standaard-Organisasie (ISO). Die unieke behoeftes van die Afrikatale met betrekking tot terminologieprosessering vind ook in samewerking met ISO (Alberts 2000a: 239) en plaaslike akademië plaas.

Die dissemineringswyse van terminologiedata in die nuwe taalbestel het ook 'n verandering ondergaan. In stede van duur tweetalige woordeboeke in 'n formele register word daar in die nuwe taalbestel probeer om korter meertalige terminologielyste met meer basiese inligting aan die terminologiegebruikers beskikbaar te stel (Ferreira en Du Plessis 2007: 73-74). Hierdie meertalige lysie is op die webblad van die Departement Kuns en Kultuur beskikbaar (<http://www.dacst.gov.za>). Disseminering deur termbanke verleen aan die gebruikers, wat sowel vakspecialiste as die algemene publiek kan insluit, vinnige, goedkoop en gerieflike toegang tot 'n wye verskeidenheid terminologiedata. In 'n poging om terminologiese aktiwiteite bymekaar te laat aansluit, het die Terminologiekoördineringsafdeling ook begin om alle afgehandelde, lopende en beplande terminologieprojekte in 'n nasionale register te registreer.

Die implementering van terminologiedata is seker een van die ingewikkelste prosesse by terminologiebestuur. Implementering van terminologiedata in die nuwe taalbestel beteken nie net die beskikbaarstelling van terminologiedata aan 'n spesifieke groep terminologiegebruikers nie, maar wel die implementering van terminologiedata aan 'n wye verskeidenheid terminologiegebruikers en die doelmatige gebruik van hierdie terminologiedata in die verskillende aktiwiteitsterreine en domeine.

Alhoewel 'n *National Language Policy Framework* (DAC 2002) en *Implementa-*

tion Plan: National Policy Framework (DAC 2003) bestaan waarvan die doelwit is om 'n samehangende taalbeleidsraamwerk en implementeringsplan vir 'n meertalige taalbestel tot stand te bring, is daar nie 'n beleid wat spesifiek op terminologie gerig is nie. Die gebrek aan 'n terminologiebeleid en implementeringsplan om terminologiedata te reguleer word as 'n leemte geïdentifiseer aangesien daar nie definitiewe riglyne en tydraamwerke bestaan waarteen die implementering van terminologiedata in Suid-Afrika gemeet kan word nie.

Die gebrek aan die bestaan van 'n terminologiebeleid is moontlik die oorsaak van die gebrek aan teksprodusering in wetenskaplike artikels in die Afrikaanse. Dit geld in 'n mate ook vir Afrikaans. Mouton (2005: 371) bevind dat die totale aantal geakkrediteerde artikels in die databasis (South African Knowledgebase) tussen 1990 en 2002, 90 713 was waarvan 81 565 (89,9%) in Engels, 7 453 (8,2%) in Afrikaans, en 1 791 (2,2%) in ander tale (hoofsaaklik Nederlands, Duits en Frans). Engels word derhalwe hoofsaaklik in wetenskaplike artikels in geakkrediteerde tydskrifte aangewend aangesien teksproduseerders skynbaar Engels verkies wat as 'n wetenskaplike en terminologiese wêreldtaal beskou word. Die feit dat terminologiedata nie deurlopend ondersteun word deur geskrewe wetenskaplike dokumentasie en tekste nie het tot gevolg dat terminologie nie deurlopend in die gemeenskap sirkuleer nie. Derhalwe word terminologie nie op 'n wyse aangebied wat vir die gebruiker 'n aanduiding gee van hoe terminologie in konteks aangewend word en hoe ander gebruikers oor die terminologie voel nie (Ferreira en Du Plessis 2007: 68).

Die evaluering van terminologie wat beskikbaar gestel word, beteken dat terminologiedata geëvalueer moet word ooreenkomstig die wyse waarop dit doelmatig deur die gebruikers aangewend word. Gebruikers van terminologie kan derhalwe op so 'n wyse betrek word dat hulle deel van die proses vorm en sodoende die kwaliteit en effektiewe gebruik van die terminologie verseker. Strategieë moet met ander woorde geïmplementeer word om terminologie en die bestuur daarvan te evalueer en die kwaliteit en sukses van die beplanning en koördinerings, navorsing, disseminering en implementering te meet. Indien die kwaliteit gemeet kan word, kan aanpassings gemaak word waar leemtes geïdentifiseer is.

Die evaluering van terminologie hang verder ook nou saam met die implementeringstrategie. Evaluering moet so plaasvind dat die resultate gebruik kan word om tot voordeel van die terminologiegebruikers te dien.

5. Slotsom

In hierdie artikel word 'n model voorgestel waarvolgens terminologiebeplanning, -koördinerings en -bestuur aangepak kan word. Wanneer die implikasies van so 'n model vir die Suid-Afrikaanse situasie verreken word, moet in gedagte gehou word dat dit 'n ideaalmodel is wat moontlik sonder enige probleme in 'n gesofistikeerde vaktamillieu geïmplementeer kan word, waar daar met ontwikkelde tale (wat hoofsaaklik in een register en op een terrein, by-

voorbeeld parlementêre terme) gewerk word. Die situasie wat terminologie en terminologie-inligting in die amptelike tale van Suid-Afrika betref, lyk egter heel anders. Terminologiebestuur in die Suid-Afrikaanse vaktaalmilieu vereis nie slegs dat terme wat in die verskillende aktiwiteitsterreine of domeine ontstaan, gedokumenteer moet word nie. Die amptelike tale (wat Afrikaans insluit) moet ook uitgebrei word om as funksionerende tale naas Engels op te tree. Terminologie-inligting word benodig vir behoorlike kommunikasie en die uitskakeling van misverstande. Daar word vanuit Engels as brontaal na tien doeltale gewerk. Die register strek vanaf skool- tot by akademiese vlak. Die aktiwiteitsterreine en domeine sluit onder meer in weerkunde, VIGS, gesondheid, dierkunde, chemie, wiskunde in die verskillende tale en verskillende registers (vlakke van begrip).

Alhoewel die implementering van die voorgestelde terminologiebestuursmodel moontlik haakplekke mag ondervind, is daar reeds gevorder rondom die uitbreiding en daarstelling van terminologiedata in al die amptelike tale van Suid-Afrika. Met die nodige terminologiebestuursleiding sal 'n terminologiebestuursprogram vir Suid-Afrika nie slegs 'n ideaal bly nie. Intendeel, dit kan gou 'n realiteit word.

Bibliografie

- Alberts, M.** 2000a. Terminology Management at the National Language Service. *Lexikos* 10: 234-251.
- Alberts, M.** 2000b. *Audit on South African Terminology Projects. July 2000.* Ongepubliseerde dokument.
- Alberts, M. en kollegas by die Terminologiekoördineringsafdeling.** 2001. Persoonlike onderhoud. Januarie 2001. Pretoria.
- Budin, G.** 1999. *Terminology, Translation, Technical Communication* [aanlyn]. <http://mailbox.univie.ac.at/~budin8/dint.htm> [15 Augustus 2008].
- Cabré, M.T.** 1999. *Terminology: Theory, Methods and Applications.* Amsterdam/Philadelphia: John Benjamins.
- Cluver, A.D. de V. en C.J. Scheffer (Reds.).** 1984. *RGN-ondersoek na vaktaalaangeleenthede in Suidelike Afrika: Verslag van die ondersoek na organisatoriese aangeleenthede op vaktaalgebied.* Term-1. Pretoria: RGN.
- Department of Arts and Culture (DAC).** 2000/2001. *Annual Report: Nasionale Taaldiens. Terminologiekoördinerings: 78-81.* Pretoria: Departement Kuns en Kultuur.
- Department of Arts and Culture (DAC).** 2001. *Business Plan: Training of Collaborators on Terminology.* Ongepubliseerde dokument.
- Department of Arts and Culture (DAC).** 2002. *National Policy Framework. Final Draft (NLPF). 13 November 2002.* Pretoria: Departement Kuns en Kultuur.
- Department of Arts and Culture (DAC).** 2003. *Implementation Plan: Framework for the National Language Policy. Final Draft. 10 April 2003.* Pretoria: Departement Kuns en Kultuur.
- Department of Arts and Culture (DAC)** 2006/2007. *Annual Report.* Pretoria: Departement Kuns en Kultuur.

- Department of Arts, Culture, Science and Technology (DACST).** 1996. *Overview, Recommendations and Executive Summary from the Final Report of the Language Plan Task Group (LANGTAG)*. Pretoria: Departement Kuns, Kultuur, Wetenskap en Tegnologie.
- Department of Arts and Culture, Science and Technology (DACST).** 1998. *Annual Report: Nasionale Taalddiens: Terminologie: 27-29*. Pretoria: Departement Kuns, Kultuur, Wetenskap en Tegnologie.
- Department of Arts and Culture, Science and Technology (DACST).** 1999. *Annual Report: Nasionale Taalddiens: Terminologie: 130-132*. Pretoria: Departement Kuns, Kultuur, Wetenskap en Tegnologie.
- Ferreira, D.M. en L.T. du Plessis.** 2007. Terminologiebestuur in Suid-Afrika. 'n Beoordeling van drie werkmodes. *Lexikos* 17: 56-76.
- Ferreira, D.M.** 2002. Terminologiebestuur in Suid-Afrika met spesifieke verwysing na die posisie van histories ingekorte tale. Ongepubliseerde Ph.D.-tesis. Bloemfontein: Universiteit van die Vrystaat.
- Galinski, C. en G. Budin.** 1993. New Trends in Translation-Oriented Terminology Management. Wright, S.E. en L.D. Wright (Reds.). 1993. *Scientific and Technical Translation: 209-215*. Amsterdam: John Benjamins.
- Grondwet.** 1996. Die Republiek van Suid-Afrika. Wet 108 van 1996.
- Haugen, E.** 1983. The Implementation of Corpus Planning: Theory and Practice. Cobarrubias, J. en J.A. Fishman (Reds.). 1983. *Progress in Language Planning. International Perspectives: 269-289*. Den Haag: Mouton.
- ISO 1024.** 1992. *International Terminology Standards — Preparation and Layout*. International Organization for Standardization.
- Mouton, J.** 2005. Afrikaans as wetenskapstaal in Suid-Afrika. *Tydskrif vir Geesteswetenskappe* 45(3): 370-385.
- Pan South African Board Act.** 1995. Act Nr. 59 of 1995.
- Pan South African Language Board Amendment Act.** 1999. Act Nr. 10 of 1999.
- Pan South African Language Board.** 2004/2005. *Annual Report*.
- Pan South African Language Board.** 2006/2007. *Annual Report*.
- Rubin, J.** 1971. Evaluation and Language Planning. Rubin, J. en B.H. Jernudd (Reds.). 1971. *Can Language be Planned?: 217-252*. Honolulu: East-West Center and University Press of Hawaii.
- Sager, J.C.** 1990. *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins.
- Wright, S.E. and G. Budin.** 1997. Introduction. Wright, S.E. and G. Budin (Reds.). 1997. *Handbook of Terminology Management: 1-12*. Amsterdam: John Benjamins.

Namibian University Entrants' Concepts of 'a Dictionary'*

Herman L. Beyer, *Department of Language and Literature Studies*
(hbeyer@unam.na)

and

André K. Faul, *Faculty of Science* (afaul@unam.na),
University of Namibia, Windhoek, Namibia

Abstract: This article reports on a survey conducted in 2009 among students registering for their first year of study at the University of Namibia. The aim of the study was to determine these school leavers' concepts of 'a dictionary' and to what extent they perceive dictionaries to be of potential help in satisfying needs for specific types of linguistic information. Furthermore, the survey attempted to determine if the variables of frequency of dictionary use, exposure to dictionary pedagogy at school and dictionary ownership had any influence on respondents' concepts of 'a dictionary'. In comparison to the literature and other similar studies, this study takes a new approach by measuring respondents' concepts of dictionaries against a pre-constructed dictionary profile based on validated assumptions while focusing on the pre-consultation situation instead of on a reflection on past dictionary consultation procedures. The findings indicate that Namibian school-leavers do not grasp the complete information potential of monolingual dictionaries. Also, although school syllabuses of language subjects require dictionary skills to be taught, about a third of the respondents were not exposed to dictionary pedagogy at school, while those respondents who were exposed to some form of dictionary pedagogy do not demonstrate a substantially different concept of dictionaries from those who did not undergo dictionary training. This result questions the quality of dictionary pedagogy where it does take place. The effects of frequency of dictionary use and dictionary ownership on respondents' concepts of dictionaries also seem to have been minimal.

Keywords: CONCEPT, DEFINITION, DICTIONARY, DICTIONARY OWNERSHIP, DICTIONARY PEDAGOGY, DICTIONARY PROFILE, DICTIONARY USER, FREQUENCY OF DICTIONARY USE, INFORMATION POTENTIAL, LEXICOGRAPHICALLY RELEVANT NEEDS, LEXICOGRAPHY, QUESTIONNAIRE, USAGE RESEARCH, USER NEEDS, USER RESEARCH

Opsomming: Namibiese universiteitstoetreders se konsepte van "'n woordeboek". Hierdie artikel doen verslag van 'n opname wat in 2009 onder eerstejaarstudente aan die Universiteit van Namibië onderneem is. Die doel van die studie was om die skoolverlaters se

* A preliminary version of this article was presented as a paper under the title "'What is a Dictionary?" Namibian University Entrants' Perceptions' at the Fourteenth International Conference of the African Association for Lexicography, organised by the Xhosa Department of the University of the Western Cape, Bellville, South Africa, 6–8 July 2009.

konsepte van " 'n woordeboek" vas te stel en te bepaal in welke mate hulle woordeboeke beskou as potensiele hulp in die bevrediging van spesifieke tipes linguistiese inligtingsbehoefes. Daarbenevens het die opname geprobeer om te bepaal of die veranderlikes van frekwensie van woordeboekgebruik, blootstelling aan woordeboekonderrig en woordeboekeienaarskap enige invloed op die respondente se konsepte van " 'n woordeboek" gehad het. In vergelyking met die literatuur en soortgelyke opnames volg hierdie studie 'n nuwe benadering deur respondente se konsepte van woordeboeke te meet teen 'n voorafgekonstrueerde woordeboekprofiel wat op geldige veronderstellings gebaseer is terwyl die fokus op die voornaslaansituasie val in plaas van op nadenke oor vorige woordeboeknaslaanprosedures. Die bevindinge toon dat Namibiese skoolverlaters nie die volledige inligtingspotensiaal van verklarende woordeboeke besef nie. Alhoewel skoolsillabusse van taalvakke vereis dat woordeboekvaardighede onderrig word, het ongeveer 'n derde van die respondente ook nie enige vorm van woordeboekonderrig ondergaan nie, terwyl dié respondente wat wel een of ander vorm van woordeboekonderrig ondergaan het, nie 'n beduidend andersoortige opvatting van woordeboeke getoon het as dié wat nie woordeboekonderrig ontvang het nie. Hierdie resultaat bevraagteken die kwaliteit van woordeboekonderrig waar dit wel plaasvind. Die effek van frekwensie van woordeboekgebruik en woordeboekeienaarskap op respondente se konsepte van woordeboeke blyk ook minimaal te wees.

Sleutelwoorde: DEFINISIE, FREKWENSIE VAN WOORDEBOEKGEBRUIK, GEBRUIKERSBEHOEFES, GEBRUIKERSNAVORSING, GEBRUIKSNAVORSING, INLIGTINGSPOTENSIAAL, KONSEP, LEKSIKOGRAFIE, LEKSIKOGRAFIES RELEVANTE BEHOEFES, VRAELYS, WOORDEBOEK, WOORDEBOEKEIENAARSKAP, WOORDEBOEKGEBRUIKER, WOORDEBOEKONDERRIG, WOORDEBOEKPROFIEL

Knowledge is of two kinds. We know a subject ourselves,
or we know where we can find information upon it.

Samuel Johnson, 1775

1. Introduction

One of the noticeable features of developments in theoretical lexicography during the last two to three decades has been a strong focus on the information needs and reference skills of the target users of dictionaries (cf. e.g. Gouws and Prinsloo 2005: 5). This focus presupposes that any dictionary is compiled for a specific purpose, and that the contents and presentation of the dictionary should be directed at that specific target group (cf. e.g. Gouws 1989: 49, Hartmann 1989). This notion has led to the study of various features of dictionaries, such as dictionary functions and structures, all of which are informed by the user and the usage situation as central components of the genuine purpose of a dictionary (cf. e.g. Gouws and Prinsloo 2005: 1-8).

This article reports on a survey conducted in 2009 among students registering for their first year of study at the University of Namibia (UNAM). The aim of the study was to determine these school leavers' concepts of 'a diction-

ary' and to what extent they perceive dictionaries to be of potential help in satisfying needs for specific types of linguistic information.

2. Positioning and relevance

2.1 User research

2.1.1 State of the art

Atkins and Rundell (2008: 30) describe the concept of dictionary user research as follows:

'User research' refers to any method used for finding out what people do when they consult their dictionaries, what they like and dislike about them, and what kinds of problem they look the dictionary to solve.

According to Tarp (2009: 276) interest in dictionary user research was stimulated by the conference on lexicography convened at Indiana University in November 1960 which resulted in the publication of Householder and Saporta (1967). Subsequently, Wiegand (1977) proposed the determination of a sociology of the dictionary user, to which Hartmann (1989) responded with twelve hypotheses about the dictionary user. Hartmann (1989) also reviewed the empirical user research reported up to that time in addressing his hypotheses, referring to, among others, the famous study of Barnhart (1967) and those of Quirk (1973), Béjoint (1981) and Greenbaum, Meyer and Taylor (1984), all of whose results indicate that monolingual dictionaries are first and foremost used for determining or confirming mostly meaning and secondly spelling. Since then, empirical user research has generally followed this line of enquiry and has been conducted on the basis of self-evaluation by users through the use of questionnaires, asking respondents to indicate frequency of use, information categories sought, reasons for dictionary use and the extent to which the consulted dictionaries have satisfied their needs (Jackson 2002: 76, Bogaards 2003: 26, Tarp 2009: 284-285). Recent examples of this type of empirical studies include Siegel (2007), and Ekwa Ebanéga and Moussavou (2008). Other instruments of empirical research employed to a lesser extent in user research are interviews, observation, protocols, experiments, tests and more recently log file analyses (Hartmann 2001: 115-120, Tarp 2009: 283-290).

While most empirical user research has focused on user behaviour during the actual dictionary consultation procedure, and has surveyed this behaviour mostly post hoc, Nesi (in Hartmann 2001: 117) has identified six 'stages' of dictionary consultation, each requiring different sets of skills from the user. These stages, which can be empirically investigated, are:

- (a) 'before study', e.g. knowing which dictionaries exist and what information can be found in them;

- (b) 'before dictionary consultation', e.g. deciding whether consultation is necessary, which might be the problem word to be searched, and which dictionary is most likely to help;
- (c) 'locating entry information', e.g. understanding the macrostructure of the dictionary (including electronic ones) and locating the problem words;
- (d) 'interpreting entry information', e.g. understanding the microstructure of the dictionary, interpreting the various information categories offered within entries (including translation equivalents in bilingual dictionaries) and extracting the relevant item(s);
- (e) 'recording entry information', e.g. deciding how to select, use and file the information extracted; and
- (f) 'understanding lexicographical issues', e.g. awareness of dictionary terminology and conventions and the ability to evaluate entries and dictionaries.

For reviews of user research up to the present day, compare Hartmann (2001: 115-120), Atkins and Varantola (2008) and Tarp (2009).

2.1.2 Criticism

Substantial and valid criticism has been levelled at empirical dictionary user research. In relation to the most popular research instrument, i.e. the questionnaire, the criticism of Hatherall (1984: 184) that the data so generated provide an indication of respondents' behaviour during questionnaire administration rather than during dictionary consultation and that 'the only reliable method of collecting data on dictionary user behaviour is by direct observation,' is the most widely cited. This criticism is confirmed by Bergenholtz and Tarp (1995: 80), who state that 'it cannot be ruled out that the problems, behaviours, etc. described by the informants differ from their real problems, etc. '; Jackson (2002: 76) mentions 'a tendency to overstate and underplay' during self-reporting; and Bogaards (2003: 26) asserts that 'it is well known that what people really do may be a far cry from what they do when interviewed.' In the field of social research, from which dictionary user research seems to derive its methods, this potential skewing of the truth is also recognised (cf. e.g. Babbie and Mouton 2001: 263). However, Lew (2002) points out that this problem of the observer's paradox is not limited to questionnaires, but pertains to all direct observation techniques. Therefore, much of the criticism also applies to other instruments, all of which have their limitations.

Nevertheless, there are at least three inherent problems with surveys asking respondents to indicate which types of information they usually seek in a dictionary with a view to determine users' needs in order to improve dictionaries. Firstly, respondents have rarely undergone dictionary pedagogy and are

at most only partially dictionary literate, which means that they are not aware of all the information categories that are at their disposal in a given dictionary (cf. Svensén 1993: 17, Hartmann 2001: 25). Siegel (2007: 30) states of the respondents in one such survey that dictionary skills had not been taught at the university where the respondents were surveyed, 'so the students would have to be learning these skills on their own'. It is then not surprising that conclusions like 'users were far from taking extensive advantage of the range of information available to them' (Algeo 1989: 31) are drawn. Secondly, and flowing from the first problem, respondents have been conditioned by previous dictionary consultation experiences to expect certain information categories in dictionaries and not to expect other information categories, leading them to consult dictionaries only when they know they would find the required information category. Consequently, conclusions such as 'the users need exactly what has already been included in dictionaries' (Tarp 2009: 291) will logically be drawn. It is therefore also understandable that Whitcut (1989: 92) reports that 'most respondents reply by describing what in their experience a dictionary is, rather than suggesting how it might be improved'. Thirdly, and flowing from the first two problems, respondents are unaware that certain existing information needs could potentially be satisfied by a dictionary if the specific information category were included in a dictionary, and consequently do not mention such needs when they are surveyed on dictionary use. In terms of Tarp (2008) these types of needs are *non-recognised lexicographically relevant potential user needs*.

Tarp (2009: 290-292) discusses 'a number of general problems with regard to the majority of lexicographical research projects published until now,' i.e. problems relating to sampling, the formulation of questions and data validation, and, specifically with regard to research on user behaviour during the actual dictionary consultation procedure, argues that any such survey should be conducted during the actual consultation procedures rather than post hoc. Tarp (2009: 283-290) also critically evaluates each type of instrument employed in empirical user research, the questionnaire instrument being the first to be categorically dismissed. However, Lew (2002) convincingly defends the questionnaire as a viable instrument, but stresses that its use should be well planned. This requirement obviously applies to other instruments of data collection as well.

2.2 The present study

2.2.1 Background

Any person who has access to a dictionary can be regarded as a potential dictionary user. In the modern theory of lexicographical functions (cf., e.g. Tarp 2008 — henceforth the 'Function Theory'), a distinction is made between the *potential dictionary user* and the *actual dictionary user*. These and other relevant concepts can be explained in a simplified way as follows: The potential diction-

ary user finds him-/herself in a so-called *extra-lexicographical situation*, i.e. a situation wherein a dictionary is not consulted. Within this situation the potential user develops various information needs, some of which may be satisfied by consulting a dictionary. Such needs are *lexicographically relevant needs*, while information needs that cannot be satisfied by consulting a dictionary are *lexicographically non-relevant needs*. When the potential user realises that a specific need is lexicographically relevant and he/she has access to an appropriate dictionary, he/she could approach the dictionary. The moment when the potential user first selects the dictionary, he/she becomes an *actual dictionary user* and enters the *dictionary usage situation*. Within this situation the dictionary consultation procedure takes place, and when the actual user has completed the consultation and shelves the dictionary or closes the application on his/her computer, he/she leaves the dictionary usage situation and returns to the status of potential user, regardless of whether the specific user need(s) were satisfied in the lexicographical situation. (Cf. Tarp 2008: 39ff.)

One tipping point between remaining a potential dictionary user and becoming an actual dictionary user by initiating the dictionary consultation procedure is the potential user's realisation or non-realisation that a specific information need is lexicographically relevant. This implies that lexicographically relevant needs could be *recognised* or *non-recognised* (Tarp 2009: 281).

To be in a position to recognise a specific information need as lexicographically relevant, a potential user must have a certain concept of what a dictionary in general is and which information categories it can contain. If a potential user has not undergone dictionary pedagogy, recognising needs as lexicographically relevant is usually based on conditioning resulting from experience in previous dictionary usage situations (cf. 2.1.2). In certain extra-lexicographical situations, a dictionary itself could remind or even bring potential users to the insight that a particular information need is lexicographically relevant, e.g. when an integrated e-dictionary on a computer produces pop-ups while the potential user is working in a word-processing application. Obviously a printed dictionary does not have these capabilities.

Consequently, Tarp (2009: 292-293) argues that empirical research aiming at identifying user needs should focus on the extra-lexicographical situation rather than on the dictionary usage situation, and that no known user research has done exactly this.

The present study, then, seems to be the first instance of user research to indeed focus on the extra-lexicographical situation, although it is not conducted within the framework of the Function Theory (but utilises some of its terms), nor is it aimed at identifying user needs per se.

2.2.2 Motivation, aims and objectives

2.2.2.1 Motivation

Information is required to contribute to:

- (a) a better understanding of potential dictionary users' lexicographical frame of reference, described by Beyer (2006: 56-57) as relating to users' assumptions, habits, perceptions and attitudes with regard to dictionaries and dictionary use;
- (b) the testing of assumptions with regard to a perceived lack of a dictionary culture;
- (c) a more precise indication than currently exists of the extent and influence of dictionary pedagogy in Namibian schools;
- (d) the establishment of a baseline for dictionary pedagogy at UNAM; and
- (e) contextual knowledge with a view to planning and promoting dictionary-integrated (language) learning in Namibian schools (cf. Beyer 2010).

With regard to motivation (c), the following learning objective appears in the senior primary phase syllabus (grades 5–7) for English as a Second Language of the Ministry of Education in Namibia (Ministry of Basic Education and Culture 2000: 10-11), which also serves as the generic syllabus for all senior primary language subject syllabuses in Namibian government schools: '[Learners will] use reference materials with speed and accuracy, e.g. dictionary, atlas, directory, encyclopaedia' [sic]. The following basic competencies are listed under this learning objective:

Learners should be able to:

- arrange words alphabetically
- use a dictionary effectively to:
 - find words
 - use headwords (e.g. happy) to find derivatives (e.g. happiness, happily)
 - find the correct spelling of a word
 - find the appropriate meaning of a word
 - find the word class of a word (n-noun, v-verb, adj-adjective)
 - recognise idiomatic phrases and expressions
 - find the correct page
 - find information

In the lower primary phase syllabus (grades 1–3), no mention is made of dictionaries, although it is stated that learners should have access to 'a wide range of readers and other reading material' (Ministry of Basic Education, Sport and Culture 2004: 2). The English Second Language Syllabus for the junior secondary phase (grades 8–10) includes the following two basic competencies under learning objectives pertaining to reading, and grammar and usage respectively (Ministry of Education 2006: 11, 16):

[Learners should be able to] demonstrate the ability to find information such as meaning, spelling, idiomatic usage in dictionaries.

[Learners should be able to] use dictionaries and other available resource books for spelling, meaning and additional information like abbreviations, idiomatic expressions, parts of speech, etc.

In the English as a Second Language Syllabus for the senior secondary phase (grades 11 and 12), a single competency statement reads that learners should be able to consult 'a dictionary for a variety of purposes' (Ministry of Education 2005: 8).

The inclusion of the respective competency statements in the syllabuses can be regarded as a ministerial directive that dictionary pedagogy should take place in Namibian government schools. This implies that school leavers should demonstrate a more than rudimentary knowledge of dictionaries as containers of linguistic information, perhaps especially those school leavers who have obtained university admission.

2.2.2.2 Aims

The aims of this study, flowing from the motivation, are to provide answers to the following research questions:

- A What are Namibian university entrants' concepts of *a dictionary*?
- B How many Namibian university entrants have been exposed to some form of dictionary pedagogy?
- C Does the undergoing of some form of dictionary pedagogy, frequency of dictionary use and ownership of dictionaries have an influence on Namibian university entrants' perceptions of what a dictionary is?

2.2.2.3 Objectives

The objectives of the survey, flowing from the aims of the study, are:

- (a) to determine how respondents would define the concept of *a dictionary*;
- (b) to determine if respondents possess rudimentary knowledge of dictionary typology;
- (c) to determine the proportion of respondents that have been exposed to dictionary pedagogy in any form;
- (d) to determine how frequently respondents used dictionaries while at school;
- (e) to determine how many respondents own dictionaries;
- (f) to determine respondents' recognised lexicographically relevant needs against a pre-constructed profile of lexicographically relevant needs; and

- (g) to determine if the variables contained in determinations (c), (d) and (e) have any influence on respondents' recognised lexicographically relevant needs as determined in (f).

The achievement of objectives (a), (b) and (f) would provide an answer to research question A. Research question B would be answered by achieving objective (c), while research question C would be answered by achieving objectives (e), (f) and (g). These objectives do not pertain to users' behaviour during the dictionary consultation procedure and therefore the survey could be conducted in the extra-lexicographical situation without compromising the results.

3. Design and execution

3.1 Population and sample

3.1.1 The population

The population (N) consists of (a) students who registered for their first year of study at UNAM in 2009, irrespective of the programme they were registering for, and (b) who completed their school-leaving examination in a school in Namibia. According to the Office of the Registrar at UNAM¹, 4 023 first year students registered at UNAM in 2009, but the computerised information system could not directly determine how many of these students, whether Namibian or not, completed their secondary education at a Namibian government school. ² To make an accurate discrimination of this nature, each first year student's record would have to be consulted. Given the sampling method employed (cf. 3.1.2.1), these steps were deemed unnecessary and therefore the exact size of the population as a subset of the 4 023 first year students at UNAM in 2009 is unknown, but does not exceed 4 023 individuals ($N < 4\ 023$).

3.1.2 The sample

3.1.2.1 Sampling method

Because of financial and time constraints, and because the members of the population are generally relatively difficult to access, a non-probability sampling method (cf. Babbie and Mouton 2001: 166) was selected for the survey. There are very few occasions when all the members of the population would be both easily accessible and willing to participate in a survey. Probably the best of these occasions would be, as is shown by Faul (2008), the registration period at the beginning of the academic year in February. At UNAM the registration procedure is largely a manual one, requiring all registering students to present themselves at the registration venue on the pre-announced registration days. During this event long queues are common and students have to wait for peri-

ods sometimes exceeding an hour before moving to the next point, during which time they would not usually resist some form of distraction without the risk of losing their place in the queue. When specifically first year students register, there is the added advantage that, because it is their first encounter with the unknown environment of a university, they are generally cooperative.

3.1.2.2 Sample type

Given the considerations above, it was decided that a *convenience sample* (cf. Du Plooy 2009: 123) would be drawn from the population: Every first year student in a registration queue willing to participate in the survey would be included. Each respondent would be asked to indicate where they completed their secondary education, but discrimination would only take place during data analysis after completion of the survey, i.e. non-members of the population also participated in the survey but were disqualified post hoc and their data excluded from the study.

3.2 Data collection method

3.2.1 The instrument

Given the sample and the circumstances under which the survey were to be conducted, the self-administered questionnaire was selected as data collection instrument.

3.2.1.1 Questionnaire design

The questionnaire (cf. Addendum A) consists of direct, specific, open-ended and closed-ended questions. All closed-ended questions were multiple-choice items, some of which allowed more than one choice to be selected. Respondents were requested to indicate their choice by making an X in the provided space(s) next to the selected option(s) (cf. Babbie and Mouton 2001: 233-243, Du Plooy 2009: 152-157).

The questionnaire contains 39 items. The first item was to be completed by the survey administrator and the remaining 38 items were to be completed by each respondent.

The first three items for the respondent requested biographical data: age, gender and the educational region in Namibia in which the respondents completed their secondary education. The following nine items requested data on what the respondents think a dictionary is, frequency of use at school, type of dictionary used, whether the respondents underwent dictionary pedagogy, and dictionary ownership. The remaining 27 items simulate potential dictionary consultation situations, each item stating an information need experienced by a hypothetical first year student named John and asking the respondents if they

think John's need would be satisfied by consulting a dictionary. The data generated from these 27 items would be used to construct the respondents' profile of recognised lexicographically relevant needs, which would be compared to a pre-constructed profile of the recognised lexicographically relevant needs of a (hypothetical) dictionary literate potential user.

The questionnaire was pre-coded to facilitate computer processing, i.e. every answer in open-ended items and each variable in multiple-choice items were assigned a numeral. Post-coding would be applied in the analysis of responses to the open-ended item requesting respondents to indicate what they think a dictionary is (item 5).

3.2.1.2 Administration

The survey period ran concurrently to the 2009 first year registration period at UNAM, i.e. from 2 to 6 February 2009. Two student assistants were recruited and trained to approach potential respondents during the registration period and, upon agreement to participate, ask the respondents in which faculty they were registering. The assistants would enter this data at item 1³ and then explain the aim and working of the questionnaire to the respondents, emphasising anonymity and the fact that there are no 'wrong' answers. The respondents were then left to complete the questionnaire on their own, but they could call the assistants if any doubt with regard to an item arose. After the questionnaires had been completed, the assistants collected them and thanked the respondents for their participation.

3.2.1.3 Completion rate

During the registration period, 650 completed questionnaires were collected. Of these, 141 questionnaires were disqualified post hoc based on the item requesting the place where the respondents had completed their secondary education, leaving 509 qualifying completed questionnaires of respondents who had indicated that they had completed secondary education in Namibia (n = 509).

This sample size represents 12.7% of the 4 023 students who registered for their first year of studies at UNAM in 2009.

3.2.1.4 Methods of data processing and analysis

The coding of questionnaire items was used to feed data into a Microsoft Excel® spreadsheet for analysis. For the responses to the open-ended question which requests respondents to provide a description of what they think a dictionary is (item 5), post-coding was applied by evaluating responses as positive or negative against predetermined criteria.

The chi-square test was employed to determine statistical significance where data sets are compared.

4. Analysis and interpretation

4.1 Composition of the sample

The sample consisted of 509 respondents (n), of which 485 indicated their gender: 175 (36.1%) male and 310 (63.9%) female. 486 respondents indicated their age, of which the average is 21.1 years. The distribution of the respondents across the thirteen educational regions of Namibia and the faculties at UNAM are not relevant to this particular study, though the data may become useful for comparison purposes should the study be replicated or students be grouped according to faculty in UNAM's compulsory academic literacy courses (relating to the aim of contributing to establishing a baseline for dictionary pedagogy).

To prevent language bias in their responses, the questionnaire did not ask of respondents to indicate their mother tongue.

4.2 Respondents' concept of 'a dictionary'

Van Sterkenburg (2003: 3) contends:

Looking for a definition of 'dictionary' is looking for a definition of the prototypical dictionary. The prototypical dictionary is the alphabetical monolingual general-purpose dictionary.

This contention is confirmed by Landau (2001: 6). Item 5 requests respondents to write in their own words what they think a dictionary is. The responses to this item would (a) indicate whether the respondents' concept of 'a dictionary' coincides with the definition of the prototypical dictionary, and (b) serve to validate the assumptions on which the rest of the survey is based.

502 respondents answered item 5 (n=502). Each response was evaluated against two criteria that were identified as conforming to the definition of the prototypical dictionary: If a respondent's definition included reference to (a) 'words' and (b) 'their meaning' in the appropriate relation, that definition was evaluated positively, i.e. conforming to the definition of the prototypical dictionary. In this way, a definition like 'a book consisting of words and their meanings' would be evaluated positively, whereas a definition like 'a book containing words and information about them' would be evaluated negatively, even though it represents a more accurate generic definition of a dictionary. Measured against these criteria, 403 (80.6%) of the 502 responses were evaluated positively. This finding supports Van Sterkenburg's definition of the prototypical dictionary and validates it as the basis for the rest of the survey.

4.3 Frequency of dictionary use

The item requesting respondents to indicate frequency of dictionary use at school (item 6) was answered by 500 respondents. Table 1 provides the data.

Table 1: Frequency of dictionary use (item 6) (n = 500; $X^2 = 97.6$; df = 3; $p < 0.05$)

| Frequency of dictionary use | % of n |
|-----------------------------|--------|
| Almost every day | 38.4 |
| Every week | 32.8 |
| Once or twice a month | 17.8 |
| Hardly ever | 11.0 |

The most significant difference exists between the number of respondents who indicated daily dictionary use (38.4%) and those who indicated that they 'hardly ever' consulted a dictionary (11.0%). Added up, 71.2% of the respondents indicated that they had used a dictionary at least once a week while at school. This is obviously a very significant proportion of the sample, and might be regarded as indicative of the extent to which learners in Namibian schools experience lexicographically relevant needs every day.

Item 6 was followed up by a contingency question (item 7) requesting respondents who indicated that they 'hardly ever' used a dictionary to indicate the reason for the low frequency of use. Of the 55 respondents (11.0% of 500) who indicated 'hardly ever', 52 answered the contingency question. Table 2 provides the data.

Table 2: Reasons for 'hardly ever' consulting a dictionary (item 7) (n = 52; $X^2 = 68.8$; df = 4; $p < 0.05$)

| Reason for 'hardly ever' consulting a dictionary | No. (% of n) |
|--|--------------|
| No dictionaries were available. | 17 (32.7) |
| Dictionary use was not allowed in class. | 1 (1.9) |
| There was one or only a few dictionaries, but I was afraid everyone would think I was stupid if I asked to use a dictionary. | 1 (1.9) |
| I did not know how to use a dictionary. | 2 (3.9) |
| I did not find it necessary to use a dictionary. | 31 (60.0) |

It is significant that most respondents who had indicated 'hardly ever' using a dictionary indicated as reason that they did not consider dictionary consultation necessary. A motivation for this behaviour might be that they are less aware of their lexicographically relevant needs than those respondents who indicated higher frequencies of dictionary consultation. The unfortunate lack of resources in many Namibian schools would account for the fact that about a third of the respondents indicated that no dictionaries were available.

4.4 Knowledge of dictionary typology

Respondents' knowledge of dictionary typology was measured with item 8, asking the respondents to indicate 'what kind of dictionary' they used at school.

This item was answered by 493 respondents (n=493). Table 3 summarises the data.

Table 3: Type of dictionary usually consulted (item 8) (n=493; $X^2=613.5$; $df=2$; $p<0.05$)

| Type of dictionary usually consulted | % of n |
|--------------------------------------|--------|
| Monolingual dictionary | 85.6 |
| Bilingual dictionary | 10.2 |
| Don't know | 4.2 |

It is significant that the overwhelming majority of respondents indicated that they had used monolingual dictionaries, while a very small minority admitted to not knowing what type of dictionary they had used. The figures relating to this distinction represents a very rough and basic idea of respondents' awareness of dictionary typology, which might have been influenced by the absence of an 'other (specify)' variable in the item. It can therefore be deduced that by far most respondents seem to be aware that at least two types of dictionaries exist, i.e. monolingual and bilingual dictionaries. The most probable reasons for the significantly low use of bilingual dictionaries is that many of the 'African' languages in Namibia either have no bilingual dictionaries with English, or the ones that do exist are of such inferior quality that they are virtually useless in potential usage situations. A welcome and modern exception is *A Khoekhoegowab Dictionary with an English Khoekhoegowab Index* (Haacke and Eiseb 2002). Such dictionaries would however be used almost exclusively in the relevant language class where the actual language is studied. In all other subjects, English is the medium of instruction (from grade 4 onwards) and the use of English monolingual dictionaries especially for text reception purposes can be assumed (but would have to be confirmed by empirical research). Also, for most of the 'African' languages in Namibia no monolingual dictionaries exist as yet.

To determine which dictionaries are popular among respondents (and schools), they were asked to indicate the title of the dictionary they usually used at school (item 9). This item was answered by 459 respondents (n=459). It contained an open-ended question, but respondents were given the option to enter "don't remember". Table 4 shows those titles indicated by more than 1% of the respondents. They are given as they have been provided by respondents, although some have been interpreted and grouped with others, as is indicated by the use of brackets.

Table 4: Titles of dictionaries usually used at school by more than 1% of the respondents (item 9) (n=459)

| Dictionary 'title' | % of n |
|---------------------|--------|
| Oxford (Dictionary) | 58.0 |
| 'Don't remember' | 24.8 |

| | |
|---|-----|
| Oxford (Advanced Learner's Dictionary) | 3.5 |
| Cambridge (Advanced Learner's) (Dictionary) | 3.0 |
| Active (English) Dictionary | 2.6 |
| Oxford School Dictionary | 1.7 |
| Longman (Dictionary) | 1.5 |
| (Longman) Contemporary English Dictionary | 1.3 |
| Oxford (English) Pocket Dictionary | 1.3 |

The results in Table 4 reflect the commercial reality in Namibia with regard to dictionaries: The dictionaries of Oxford University Press (OUP) dominate the reference sections in all bookshops in Windhoek and most bookshops in other towns. It is therefore not surprising that the vast majority of respondents seem to have used some Oxford® dictionary or other, regardless of whether it is the ideal dictionary to satisfy the potential users' information needs. Consequently, the fact that most respondents have indicated that they had used an Oxford® dictionary at school should not be interpreted as reflecting the success with which these dictionaries satisfy their lexicographical needs, nor that the Namibian education authorities prefer these dictionaries above others based on lexicographical merits. Rather, it indicates the effect of OUP's marketing efforts in Namibia. Dictionary publishers in South Africa generally view the size of the Namibian market as negligible. Therefore not all publishers market their lexicographical products with equal vigour in the country. Owing to space considerations, the effects of this attitude will not be elaborated on here.

It is interesting that 24.8% of the respondents did not remember the title of the dictionary they used at school. There could be various explanations for this phenomenon, none of which, however, can be argued for conclusively within the limits of this study.

4.5 Dictionary pedagogy

Item 10 requested respondents to indicate whether they had received any form of dictionary pedagogy at school. The item was answered by 505 respondents, of which 60.8% indicated that they had been 'taught' how to use a dictionary, while 39.2% indicated that they had not undergone dictionary pedagogy ($n = 505$; $X^2 = 24.5$; $df = 1$; $p < 0.05$).

Respondents who indicated that they had been 'taught' dictionary skills were subsequently asked whether these skills were taught to them by a language teacher or another teacher (item 11). All 308 of these respondents (60.8% of 505) answered this question, of which 94.2% indicated that they had been 'taught' dictionary skills by a language teacher, while the remaining 5.8% indicated that another teacher had 'taught' them dictionary skills ($n = 308$; $X^2 = 240.2$; $df = 1$; $p < 0.05$).

It is significant that the vast majority of respondents had undergone some

form of dictionary pedagogy, and that in the vast majority of these cases the language teacher had been the instructor. This indicates a certain level of compliance with the language syllabus requirements, but leaves cause for concern about the unduly large proportion of respondents (39.2%) who were not exposed to dictionary pedagogy in spite of syllabus requirements, and therefore had to achieve the relevant competencies on their own. This may be due to a lack of resources, teachers' own lack of dictionary skills and/or their indifference or even aversion to dictionary use, as reported by, among others, Béjoint (1989: 208-209), Carstens 1995 (106-109), Bogaards (2003: 28) and Kernerman (2007: 141).

4.6 Dictionary ownership

Item 12 requested respondents to indicate whether they own a dictionary. The item was answered by 474 respondents, of whom 74.3% indicated that they do own a dictionary and the remaining 25.7% indicated that they do not ($n=474$; $X^2=111.6$; $df=1$; $p<0.05$).

It is significant that a substantially larger proportion of the respondents owns dictionaries, compared to the proportion that does not own dictionaries. At first glance this is a somewhat surprising result, as such a high level of dictionary ownership was not expected against the background of a virtually absent reading culture in Namibia (cf. Töttemeyer 2010). However, it should be borne in mind that English is the only official language in Namibia while it was spoken as mother tongue by only 0.8% of the Namibian population in 1991 (Maho 1998: 166), and that from grade 4 onwards English is phased in as the main medium of instruction in government schools until being the sole medium of instruction by grade 8 (Ministry of Education and Culture 1993: 6-7). In this context, the relatively high level of dictionary ownership can be regarded as a clear indication of the significance respondents attach to their own experiences of lexicographically relevant needs, probably with regard to English. Whether these respondents own the appropriate dictionaries for satisfying these needs, however, is a completely different question — one that this survey does not aim to answer. What becomes clear from this and the previous items, though, is that a significant level of recognised lexicographically relevant needs do exist among the respondents, and that it is probably safe to assume that these needs would almost invariably originate in educational situations.

4.7 Recognised lexicographically relevant needs

Respondents' recognised lexicographically relevant needs were tested by items 13 to 39. For every item, respondents are presented with a statement about an information need experienced by a student named John. Following each statement, respondents are asked to indicate whether they thought a dictionary

would satisfy the relevant information need by answering 'yes', 'no' or 'don't know'. A summary of the results of these items appears in Addendum B.

Among the 27 items, seven unevenly distributed items are included to which respondents' answers should be 'no' if they demonstrate an accurate understanding of what a dictionary is, i.e. if they provided at least the prototypical definition (see 4.2) of a dictionary in item 4. These items are items 14 (i.e. 'John wonders how the weather will be like tomorrow. Do you think a dictionary would provide the answer?'), 17, 21, 26, 28, 32 and 36. A summary of the results of these items (as an extract of Addendum B) appears in Table 5 below.

Table 5: Summary of results of items 14, 17, 21, 26, 28, 32 and 36

| Item no. | Yes % | No % | Don't know % | n | X ² | df | p |
|----------|-------|------|--------------|-----|----------------|----|-------|
| 14 | 1.0 | 99.0 | 0.0 | 501 | 972.3 | 2 | <0.05 |
| 17 | 1.6 | 97.4 | 1.0 | 501 | 925.5 | 2 | <0.05 |
| 21 | 1.9 | 97.5 | 0.6 | 476 | 881.5 | 2 | <0.05 |
| 26 | 1.2 | 97.1 | 1.7 | 481 | 879.8 | 2 | <0.05 |
| 28 | 11.5 | 79.5 | 9.0 | 478 | 458.9 | 2 | <0.05 |
| 32 | 1.4 | 96.9 | 1.7 | 478 | 868.1 | 2 | <0.05 |
| 36 | 1.0 | 98.2 | 0.8 | 496 | 938.7 | 2 | <0.05 |

On average 95.1% of the respondents (n) answered 'no' to the items in Table 5, which can be regarded as an indication of the validity of the results of items 13 to 39.

The comparatively high number of 'yes' answers to item 28 (11.5%) is interesting, and it is conceivable that some dictionaries might in fact satisfy the information need of why Christmas falls on 25 December by including the relevant encyclopaedic data. For this survey, however, this data type is excluded as a lexicographical data type.

The remaining 20 items represent different data categories that are usually found in monolingual (learner's) dictionaries. The fact that the vast majority of the respondents' definitions of 'a dictionary' (80.6%) conform to the definition of the prototypical dictionary, confirms that this approach corresponds optimally to the most probable concept that the respondents have of a dictionary. These items will henceforth be called the *lexicographically valid items*. Table 6 below indicates the relevant data category represented by each lexicographically valid item.

Table 6: Data categories represented by the lexicographically valid items

| Item no. | Data category | Item no. | Data category |
|----------|---------------|----------|----------------------------|
| 13 | spelling | 27 | morphology: verb tense |
| 15 | meaning | 29 | meaning of an abbreviation |

| | | | |
|----|---------------------------------|----|------------------------------|
| 16 | pronunciation | 30 | syntagmatic lexical relation |
| 18 | break-off of word | 31 | meaning |
| 19 | morphology: plural formation | 33 | part of speech |
| 20 | meaning of an idiom | 34 | lexical relation: opposite |
| 22 | fixed preposition (collocation) | 35 | lexical relation: synonym |
| 23 | sociostylistic markedness | 37 | sociostylistic markedness |
| 24 | spelling: capitalisation | 38 | punctuation |
| 25 | etymology | 39 | subject-verb agreement |

The responses to these items were used to measure the respondents' recognised lexicographically relevant needs in two ways, namely (a) by calculating the average respondent score, and (b) by comparing responses to individual lexicographically valid items and the represented data categories.

4.7.1 Average respondent score

The data categories represented by the lexicographically valid items constitute a pre-constructed dictionary profile against which the respondents' concepts of 'a dictionary' can be measured. The average respondent score is determined by calculating the average percentage of 'yes' responses to all the lexicographically valid items. An average score of 100% would mean a 100% correspondence between the pre-constructed dictionary profile and the respondents' average concepts of 'a dictionary'.

When the percentages of 'yes' responses to the lexicographically valid items in Addendum B are taken, the average respondent score is calculated as 66.5%. It can therefore be stated that, on average, the respondents' concepts of 'a dictionary' correspond at 66.5% to the pre-constructed dictionary profile. If the pre-constructed dictionary profile is accepted as a general dictionary profile, the result can be generalised to the effect that the respondents on average recognise 66.5% of the information potential of dictionaries, which represents their recognised lexicographically relevant needs relative to the general dictionary profile. Conversely, 33.5% of the information potential of dictionaries is not recognised by the respondents, which represents their non-recognised lexicographically relevant needs relative to the general dictionary profile.

4.7.2 Responses to individual lexicographically valid items

An overview of the responses to the lexicographically valid items, as a representation of the relevant data in Addendum B, is presented in Figure 1.

Because of space constraints only the most salient features of the results will be pointed out. The presentation of the data allows the reader to study the results in detail.

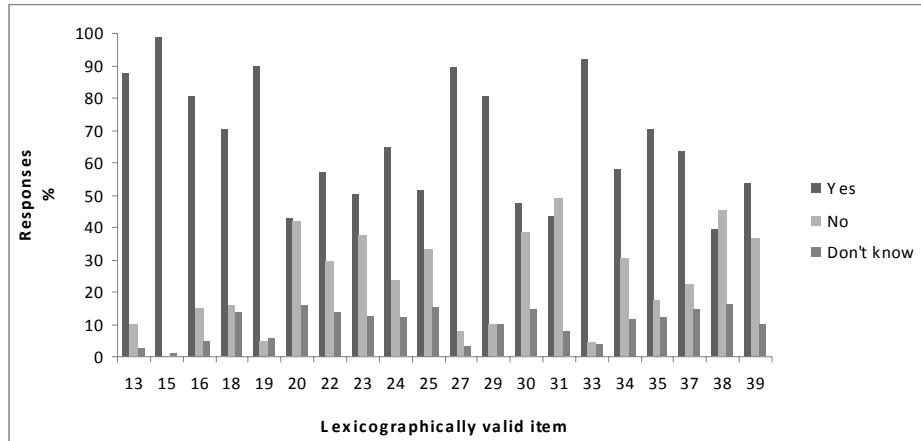


Figure 1: Bar graph showing responses to individual lexicographically valid items

The items recognised as representing lexicographically relevant needs by more than 80% of the respondents are indicated in descending order in Table 7.

Table 7: Lexicographically valid items receiving more than 80% 'yes' answers

| Position | Data category (item) | % 'yes' answers |
|----------|--|-----------------|
| 1 | meaning (15) | 98.8 |
| 2 | part of speech identification (33) | 91.8 |
| 3 | morphological information: plural formation (19) | 89.7 |
| 4 | morphological information: verb tense (27) | 89.4 |
| 5 | spelling (13) | 87.8 |
| 6 | meaning of abbreviation (29) | 80.6 |
| 7 | pronunciation (16) | 80.5 |

The items recognised as representing lexicographically relevant needs by less than 50% of the respondents are indicated in ascending order in Table 8.

Table 8: Lexicographically valid items receiving less than 50% 'yes' answers

| Position | Data category (item) | % 'yes' answers |
|----------|------------------------------------|-----------------|
| 1 | punctuation (38) | 39.1 |
| 2 | meaning of idioms (20) | 42.5 |
| 3 | meaning (31) | 43.6 |
| 4 | syntagmatic lexical relations (30) | 47.2 |

The first and most noticeable feature of the results in Table 7 and Table 8 is that the data category of meaning was recognised as lexicographically relevant by 98.8% of the respondents as represented by item 15 but only by 42.5% of the respondents as represented by item 31. Item 31 also yielded the most 'no' answers of the lexicographically valid items (43.6%). This apparent anomaly demands a closer look at the statements contained in the items involved:

Item 15: John wonders what the meaning of the word *courteous* is.

Item 31: John wonders how many eggs are in a *dozen*: 10 eggs or 12 eggs?

While the statement in item 15 can be regarded as referring directly to meaning, the statement in item 31 might not be seen by all respondents as relating to meaning. It can be argued that what is at stake in item 31 is a single semantic component of the meaning of the lexical item *dozen* and not its entire meaning. Had the statement read 'John wonders what the meaning of the word *dozen* is', the result would in all probability have compared favourably to that of item 15. Nevertheless, this anomaly provides an important insight for dictionary pedagogy: Although the need for information on the meaning of a word is clearly the most widely recognised lexicographically relevant need, as this study and those referred to in 2.1.1 confirm, the concept *meaning* and its application cannot be assumed to be identical for all potential dictionary users. During dictionary pedagogy it might not be sufficient to merely state that a dictionary (also) provides the meaning of words and to task learners during related exercises to 'find the meaning' of a given word. The full scope of the concept *meaning* should be explored, so that potential dictionary users would in the first place recognise a relevant need for information as one pertaining to (an aspect of) *meaning*, failing which, they would not recognise such a need as lexicographically relevant, as the results of item 31 seem to suggest.

The second noticeable feature of the results in Table 7 is that the data categories of part of speech identification and morphological information on plural formation and verb tense scored higher as recognised lexicographically relevant needs than spelling, which other user studies have ranked as the second most consulted data category (cf. Hartmann 1989). When the dictionary 'titles' provided by respondents in answering item 9 (cf. 4.4) are considered, the most probable reason for this difference is that the respondents used dictionaries to satisfy mostly second language information needs whereas the results referred to by Hartmann pertain to mother tongue information needs.

The items that were recognised as representing lexicographically relevant needs by between 50% and 80% of the respondents were items 18 (break-off of a word), 22 (indication of a fixed preposition), 23 (sociostylistic markedness), 24 (spelling: capitalisation), 25 (etymology), 34 (indication of an opposite), 35 (indication of a synonym), 37 (sociostylistic markedness) and 39 (subject-verb agreement).

4.8 Recognised lexicographically relevant needs vs. exposure to dictionary pedagogy at school, frequency of dictionary use or dictionary ownership

To determine if the variables of exposure to dictionary pedagogy at school, frequency of dictionary use or dictionary ownership had an effect on respondents' recognised lexicographically relevant needs relative to the pre-constructed dictionary profile, the results of two calculations are presented: (a) comparisons between the relevant average respondent scores (cf. 4.7.1) and (b) comparisons between the responses to the lexicographically valid items.

4.8.1 Compared average respondent scores for all variables

No statistically significant difference was found between the average respondent scores for any of the variables. This means that the variables of exposure to dictionary pedagogy, frequency of dictionary use and dictionary ownership had no significant effect on respondents' average scores in relation to the pre-constructed dictionary profile.

4.8.2 Responses to lexicographically valid items vs. exposure to dictionary pedagogy

The results of this comparison show that there are statistically significant differences between the two groups of responses only with regard to five of the twenty lexicographically valid items ($p < 0.05$). These results are summarised in Table 9.

Table 9: Lexicographically valid items yielding statistically significant differences between the answers of respondents who had undergone dictionary pedagogy at school and those who had not

| Item no. | Taught? | Yes % | No % | Don't know % | n | X ² | df | p |
|----------|---------|-------|------|--------------|-----|----------------|----|------|
| 13 | Yes | 90.8 | 6.9 | 2.3 | 305 | 7.2 | 2 | 0.03 |
| | No | 83.0 | 13.9 | 3.1 | 194 | | | |
| 22 | Yes | 61.8 | 25.1 | 13.1 | 283 | 8.2 | 2 | 0.02 |
| | No | 48.9 | 36.0 | 15.1 | 186 | | | |
| 24 | Yes | 69.4 | 19.4 | 11.2 | 284 | 7.9 | 2 | 0.02 |
| | No | 57.3 | 29.7 | 13.0 | 185 | | | |
| 29 | Yes | 80.8 | 7.7 | 11.5 | 287 | 6.8 | 2 | 0.03 |
| | No | 80.1 | 13.4 | 6.5 | 186 | | | |
| 34 | Yes | 59.3 | 26.8 | 13.9 | 287 | 6.5 | 2 | 0.04 |
| | No | 55.1 | 36.4 | 8.5 | 187 | | | |

The represented data categories are spelling (items 13 and 24), the indication of a fixed preposition (collocation) (item 22), the meaning of an abbreviation (item 29) and the indication of an opposite (item 28). With regard to the items pertaining to spelling, the statistical significant difference occurs with regard to the number of 'yes' answers, i.e. significantly more respondents who had undergone dictionary pedagogy at school recognised spelling as a lexicographically relevant need than those who had not been exposed to dictionary pedagogy. The same applies to the item representing the indication of a fixed preposition. With regard to the items representing an indication of the meaning of an abbreviation and the indication of an opposite, the statistically significant difference occurs in relation to the number of 'no' answers, i.e. significantly more respondents who had not undergone dictionary pedagogy denied these data categories as lexicographically relevant needs than those who had indeed been exposed to dictionary pedagogy.

Overall, exposure to dictionary pedagogy seems to have had a significant effect on respondents' evaluation of specific information needs as lexicographically relevant with regard to five (25%) of the twenty lexicographically valid items. These recognised lexicographically relevant needs correspond to a certain extent to the competency statements in the relevant school syllabuses (cf. 2.2.2.1).

4.9 Responses to lexicographically valid items vs. frequency of dictionary use

When responses to lexicographically valid items according to frequency of dictionary use are compared, six of the lexicographically valid items yielded significantly different results ($p < 0.05$). These results are summarised in Table 10.

Table 10: Lexicographically valid items yielding statistically significant differences in terms of frequency of dictionary use

| Item no. | Frequency of use | Yes % | No % | Don't know % | n | X ² | df | p |
|----------|------------------|-------|------|--------------|-----|----------------|----|-------|
| 15 | Daily | 99.5 | 0.0 | 0.5 | 189 | 17.1 | 6 | 0.01 |
| | Weekly | 100.0 | 0.0 | 0.0 | 162 | | | |
| | Monthly | 100.0 | 0.0 | 0.0 | 88 | | | |
| | Rarely | 90.7 | 1.9 | 7.4 | 54 | | | |
| 19 | Daily | 91.5 | 5.1 | 3.4 | 177 | 17.3 | 6 | 0.01 |
| | Weekly | 91.6 | 2.6 | 5.8 | 154 | | | |
| | Monthly | 86.2 | 9.2 | 4.6 | 87 | | | |
| | Rarely | 82.4 | 2.0 | 15.6 | 51 | | | |
| 22 | Daily | 65.6 | 23.7 | 10.7 | 177 | 14.1 | 6 | 0.028 |
| | Weekly | 54.6 | 29.2 | 16.2 | 154 | | | |
| | Monthly | 51.2 | 38.4 | 10.4 | 86 | | | |
| | Rarely | 44.0 | 34.0 | 22.0 | 50 | | | |

| | | | | | | | | |
|----|---------|------|------|------|-----|------|---|--------|
| 29 | Daily | 85.2 | 5.5 | 9.3 | 182 | 26.1 | 6 | 0.0002 |
| | Weekly | 79.9 | 10.1 | 10.0 | 149 | | | |
| | Monthly | 85.2 | 11.4 | 3.4 | 88 | | | |
| | Rarely | 56.9 | 21.6 | 21.5 | 51 | | | |
| 33 | Daily | 95.1 | 1.6 | 3.3 | 182 | 19.9 | 6 | 0.003 |
| | Weekly | 91.9 | 4.7 | 3.4 | 148 | | | |
| | Monthly | 92.1 | 3.4 | 4.5 | 89 | | | |
| | Rarely | 78.4 | 15.7 | 5.9 | 51 | | | |
| 39 | Daily | 61.5 | 28.9 | 9.6 | 187 | 14.5 | 6 | 0.02 |
| | Weekly | 52.2 | 41.4 | 6.4 | 157 | | | |
| | Monthly | 46.6 | 39.8 | 13.6 | 88 | | | |
| | Rarely | 40.8 | 44.4 | 14.8 | 54 | | | |

The two items yielding the greatest statistically significant differences are item 29 ($p=0.0002$) and item 33 ($p=0.003$). These items represent the data categories of meaning of an abbreviation and part of speech identification respectively. In terms of item 29, significantly fewer respondents who indicated that they had rarely used a dictionary at school answered 'yes' and significantly more 'no' and 'don't know' than those who indicated more frequent dictionary use. This leads to the conclusion that significantly more respondents who had used dictionaries at least once a month recognised the need for information on the meaning of an abbreviation as a lexicographically relevant need than those who had rarely used a dictionary. A similar result applies to item 33 with respect to recognising the need for information on part of speech as lexicographically relevant, except that the differences are only significant in terms of the 'yes' and 'no' answers.

The results of item 15 indicate that significantly more respondents who had used a dictionary rarely were unsure of whether needs for information on the meaning of words were lexicographically relevant than those who had more frequently used a dictionary and answered 'don't know' (7.4% as opposed to 0.5% and lower). Also, significantly fewer respondents who indicated rare dictionary use confirmed this need as lexicographically relevant by answering 'yes' (90.7% as opposed to 99.5% and higher), although the absolute result remains very high.

Significantly more respondents who indicated rare dictionary use were unsure whether a need for information on the formation of plural forms represents a lexicographically relevant need and answered 'don't know' than those who indicated more frequent dictionary use (item 19: 15.6% as opposed to 5.8% and lower).

With regard to item 22, significantly more respondents who indicated daily dictionary use confirmed a need for information on a collocation as lexicographically relevant than those who indicated less frequent dictionary use (65.6% as opposed to 54.6% and lower). Also, significantly more respondents who indicated rare dictionary use were unsure of the status of this information

need than those who indicated more frequent dictionary use (22.0% as opposed to 16.2% and lower). Generally, however, the need for information on collocations are regarded as lexicographically relevant by only 65.6% of respondents who indicated daily dictionary use, and declines with decreased frequency of dictionary use.

The results to item 39 shows a clear decrease in the number of 'yes' answers with a decrease in dictionary use frequency (daily 61.5% to rarely 40.8%), suggesting that significantly more respondents who indicated daily dictionary use recognised a need for information on subject-verb agreement as lexicographically significant than those who indicated less frequent dictionary use.

Generally, the results described above indicate that more respondents who used dictionaries frequently recognised the mentioned needs as lexicographically relevant than those who used dictionaries less frequently. This is the case with six (30%) of the twenty lexicographically valid items.

4.10 Responses to lexicographically valid items vs. dictionary ownership

In comparing responses to lexicographically valid items according to dictionary ownership, only two of the lexicographically valid items yielded significantly different results ($p < 0.05$). These results are summarised in Table 11.

Table 11: Lexicographically valid items yielding statistically significant differences in terms of dictionary ownership

| Item no. | Owner of dictionary? | Yes % | No % | Don't know % | n | X ² | df | p |
|----------|----------------------|-------|------|--------------|-----|----------------|----|-------|
| 19 | Yes | 91.8 | 3.0 | 5.2 | 330 | 9.6 | 2 | 0.008 |
| | No | 82.6 | 9.6 | 7.8 | 115 | | | |
| 37 | Yes | 66.5 | 21.5 | 12.0 | 343 | 7.8 | 2 | 0.02 |
| | No | 54.3 | 24.1 | 21.6 | 116 | | | |

In terms of item 19, significantly more respondents who own dictionaries recognised a need for information on plural formation as lexicographically relevant and answered 'yes' than did respondents who do not own dictionaries (91.8% vs. 82.6%). Conversely, significantly more respondents who indicated that they do not own a dictionary denied this need as lexicographically relevant and answered 'no' than did respondents who indicated that they own a dictionary (9.6% vs. 3.0%).

With regard to item 37, significantly more respondents who own dictionaries recognised a need for pragmatic information on a word as lexicographically relevant and answered 'yes' than did respondents who do not own dictionaries (66.5% vs. 54.3%). Also, significantly more respondents who indicated that they do not own a dictionary were unsure whether this need is lexicographically relevant and answered 'don't know' than did respondents who indicated that they own a dictionary (21.6% vs. 12.0%).

Overall, dictionary ownership as a variable seems to have had a significant effect on respondents' evaluation of specific information needs as lexicographically relevant with regard to two (10%) of the twenty lexicographically valid items.

5. Conclusions

The results discussed above have addressed the objectives of the survey as set out in 2.2.2.3. In terms of the aims of the study, the following conclusions, as answers to the research questions in 2.2.2.2, can be formulated:

- A Generally, the respondents' perceptions of 'a dictionary' coincides with a prototypical definition of the concept *dictionary* as 'a source providing the meaning of words', and in line with this perception respondents demonstrate a rudimentary knowledge of dictionary typology.
- B In spite of the syllabus requirements of the Namibian Ministry of Education, only 60.8% of the respondents had been exposed to some form of dictionary pedagogy at school.
- C (1) Dictionary pedagogy at school has had a minimal effect (25%) on respondents in terms of recognising specific needs for information in the survey as lexicographically relevant. In terms of the average concept respondents have of the information potential of a dictionary, it seems that dictionary pedagogy at school has had no effect.

(2) Frequency of dictionary use has had some effect (30%) on respondents in terms of recognising specific needs for information in the survey as lexicographically relevant. In fact, it seems that frequency of dictionary use has had a marginally greater effect in this respect than dictionary pedagogy. In terms of the average concept respondents have of the information potential of a dictionary, it seems that frequency of dictionary use has had no effect.

(3) Dictionary ownership has had a negligible effect (10%) on respondents in terms of recognising specific needs for information in the survey as lexicographically relevant. In terms of the average concept respondents have of the information potential of a dictionary, it seems that dictionary ownership has had no effect.

6. Limitations of the study

The following limitations apply to this study, and offer avenues for further research:

- (a) Because of the sample type, the results of the survey cannot be generalised for the whole population. However, the sample is fairly substantial

in size (n=509), providing at least some level of confidence in its representatives. To acquire results that could be applied to the whole population, either replications of the survey (cf. Du Plooy 2009: 122) or complementing data gathering methods would have to be employed (cf. Lew 2002).

- (b) The quality of dictionary pedagogy as a variable is not accounted for. The general assumption, however, based on the results of the survey, knowledge of teacher training curricula and experience with in-service teacher training, is that dictionary pedagogy at schools would be offered at an extremely basic level and in a non-systematic fashion. This could probably be attributed to teachers' own insecurities and attitudes, as referred to in 4.5.
- (c) Although the lexicographically valid items in the questionnaire are said to be representing data categories, the exact level of representation is not accounted for. This, however, does not influence the relative results. As was seen in the different responses to items 13 and 31, both said to represent the data category of indication of meaning, the formulation of items might have an influence on responses. This possible influence could be accounted for by including more (control) items representing a single data category.

7. Perspective

This perspective is offered as a response to the motivation for this study (cf. 2.2.2.1).

The concepts the respondents have of 'a dictionary' resemble the concept of the prototypical dictionary.

Carstens (1995: 107) states that school learners display an aversion to dictionary use. This observation seems to be challenged by the fact that 38.4% of the respondents in this study reported daily dictionary use and a further 32.8% weekly dictionary use at school, while 74.3% of the respondents own dictionaries. One reason for the apparent inconsistency between Carstens' statement and the data could be that, while Carstens' statement applies to the South African context, the educational context in Namibia causes comparatively more (acute) needs for linguistic information among learners. Another, perhaps less probable reason could be that learners' attitudes to dictionaries and dictionary use have changed significantly over the past fifteen years.

It is generally accepted that no dictionary culture exists in South Africa, and the same can probably be said of Namibia. However, when the frequency of dictionary use and the levels of dictionary ownership revealed by the survey are considered, at least some prerequisites for the establishment of a dictionary culture seems to exist among the respondents, notably the *self*-recognition of lexicographically relevant needs. If the results of this study could be general-

ised for the population of school learners in Namibia, then the country is certainly ripe for the organised implementation of a dictionary culture by employing a broad framework like that of Klein (2007).

One of the focal areas for the implementation of a dictionary culture would indeed be at schools. At the moment, however, dictionary pedagogy in Namibian schools seems to be largely ineffective. This could probably be attributed to the lack of a reading culture prevailing in the country and not in the least among school teachers. Language teachers themselves have generally not been exposed to dictionary pedagogy either. It is therefore highly probable that the teaching of dictionary skills is grounded in the same deficient concepts of dictionaries that this study has revealed. (Re-)implementing dictionary pedagogy at schools would have to start with (a) in-service teacher training, (b) the inclusion of dictionary pedagogy in student teacher curricula, and (c) more detailed competency statements in the relevant school syllabuses.

Finally, as a result of the above, this study has shown that the planning of dictionary pedagogy at UNAM should assume no prior exposure to (effective) dictionary pedagogy, while the notion of *meaning* should not be assumed to be identical for all potential users and that dictionary pedagogy should take cognisance of this important aspect.

Endnotes

1. The authors are grateful to Ms Annelie van der Hoeven, Assistant Registrar: Academic Administration, for her assistance.
2. The system could determine that 978 first year students had completed their secondary education at a Namibian government school in 2008, but not all students who registered for their first year in 2009 necessarily completed their secondary education in 2008.
3. In the reproduced questionnaire in Addendum A, item numbers are indicated in italics and between brackets, e.g. (1).

References

- Algeo, J. 1989. Dictionaries as Seen by the Educated Public in Great Britain and the USA. Hausmann, F.J. et al. (Eds.). 1989–1991: 28-34.
- Atkins, B.T.S. and K. Varantola. 2008. Monitoring Dictionary Use. Fontenelle, T. (Ed.). 2008. *Practical Lexicography. A Reader*: 337-375. Oxford: Oxford University Press.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Babbie, E. and J. Mouton (Eds.). 2001. *The Practice of Social Research*. Cape Town: Oxford University Press Southern Africa.
- Barnhart, C.L. 1967. Problems in Editing Commercial Monolingual Dictionaries. Householder, F.W. and S. Saporta. (Eds.). 1967: 161-180.
- Béjoint, H. 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2: 207-222.

- Béjoint, H.** 1989. The Teaching of Dictionary Use: Present State and Future Tasks. Hausmann, F.J. et al. (Eds.). 1989–1991: 208-215.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries.* Amsterdam/Philadelphia: John Benjamins.
- Beyer, H.L.** 2006. 'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelike in woordeboeke. *Lexikos* 16: 46-66.
- Beyer, H.L.** 2010. Kontekstuele bepalers by die beplanning van skoolwoordeboeke. *Lexikos* 20: 52-72.
- Bogaards, P.** 2003. Uses and Users of Dictionaries. Van Sterkenburg, P. (Ed.). 2003: 26-33.
- Carstens, A.** 1995. Language Teaching and Dictionary Use: An Overview. *Lexikos* 5: 105-116.
- Du Plooy, G.M.** 2009. *Communication Research. Techniques, Methods and Applications.* Second Edition. Cape Town: Juta.
- Ekwa Ebanéga, G.-M. and F. Tomba Moussavou.** 2008. A Survey of the Dictionary Use of Gabonese Students at Two South African Universities. *Lexikos* 18: 349-365.
- Faul, A.K.** 2008. Increasing Interactions with Nature: A Survey of Expectations on a University Campus. *Urban Habitats* 5(1): 58-83.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: SUN PReSS.
- Gouws, R.H.** 1989. *Leksikografie.* Pretoria/Cape Town: Academica.
- Greenbaum, S., C.F. Meyer and J. Taylor.** 1984. The Image of the Dictionary for American College Students. *Dictionaries* 6: 31-52.
- Haacke, W.H.G. and E. Eiseb.** 2002. *A Khoekhoegowab Dictionary with an English Khoekhoegowab Index.* Windhoek: Gamsberg Macmillan.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989–1991: 102-111.
- Hartmann, R.R.K.** 2001. *Teaching and Researching Lexicography.* Essex: Pearson Education.
- Hatherall, G.** 1984. Studying Dictionary Use: Some Findings and Proposals. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings:* 183-189. Tübingen: Max Niemeyer.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie.* Berlin/New York: Walter de Gruyter.
- Householder, F.W. and S. Saporta (Eds.).** 1967. *Problems in Lexicography.* Bloomington: Indiana University.
- Jackson, H.** 2002. *Lexicography. An Introduction.* London/New York: Routledge.
- Kernerman, I.J.** 2007. What's So Good or Bad about Advanced EFL Dictionaries? Gottlieb, H. and J.E. Mogensen (Eds.). 2007. *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004:* 139-145. Amsterdam/Philadelphia: John Benjamins.
- Klein, J.** 2007. Implementing a Dictionary Culture in South Africa: An Attempt at a General Framework for the African Languages. *Lexikos* 17: 416-423.
- Landau, S.I.** 2001. *Dictionaries. The Art and Craft of Lexicography.* Second Edition. Cambridge: Cambridge University Press.
- Lew, R.** 2002. Questionnaires in Dictionary Use Research: A Reexamination. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Congress, EURALEX*

- 2002, Copenhagen, Denmark, August 12–17, 2002, Vol. 1: 267-271. Copenhagen: Center for Sprogteknologi, Copenhagen University.
- Maho, J.F.** 1998. *Few People, Many Tongues. The Languages of Namibia*. Windhoek: Gamsberg Macmillan.
- Ministry of Basic Education and Culture.** 2000. *Upper Primary Phase Syllabus. Second Language Integrated English Syllabus. Grades 5–7*. Okahandja: National Institute for Educational Development.
- Ministry of Basic Education, Sport and Culture.** 2004. *The Curriculum for the Lower Primary Phase (Grades 1–4)*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2005. *Namibia Senior Secondary Certificate English Second Language Syllabus. Ordinary Level. Grades 11–12*. Okahandja: National Institute for Educational Development.
- Ministry of Education.** 2006. *Junior Secondary Phase English Second Language Syllabus. Grades 8–10*. Okahandja: National Institute for Educational Development.
- Ministry of Education and Culture.** 1993. *The Language Policy for Schools. 1992–1996 and Beyond*. Windhoek: Ministry of Education and Culture.
- Quirk, R.** 1973. The Social Impact of Dictionaries in the UK. McDavid, R.I. and A.R. Duckert (Eds.). 1973. *Lexicography in English*: 76-83. New York: New York Academy of Sciences.
- Siegel, M.E.A.** 2007. What Do You Do with a Dictionary? A Study of Undergraduate Dictionary Use. *Dictionaries* 28: 23-47.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S.** 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.
- Töttemeyer, A.** 2010. Multilingualism and the Language Policy for Namibian Schools. *Education in Fragile Contexts: Governance Practices and Challenges*. Baden-Baden: Nomos.
- Van Sterkenburg, P.** 2003. 'The' Dictionary: Definition and History. Van Sterkenburg, P. (Ed.). 2003: 3-17. Amsterdam/Philadelphia: John Benjamins.
- Van Sterkenburg, P. (Ed.).** 2003. *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins.
- Whitcut, J.** 1989. The Dictionary as a Commodity. Hausmann, F.J. et al. (Eds.). 1989–1991: 88-94.
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand. 1997. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zürich: Bibliographisches Institut.

Addendum A: The questionnaire



**The University of Namibia
Dictionary Survey**

(1)

| |
|----------|
| 1__ FE |
| 2__ FEMS |
| 3__ FEN |
| 4__ FHSS |
| 5__ FL |
| 6__ FMHS |

This is a survey to find out what you think about dictionaries. You will not write your name on this questionnaire, so we will not know who answered on this page. Please try to be as honest as possible. There are no right or wrong answers. We just want to find out what students think about dictionaries in general.

Please write only in the *white* spaces or make an X where you are requested to do so.

| | | | |
|---------------|--|-------|---|
| (2) Your age: | | years | 7 |
|---------------|--|-------|---|

| | | |
|--|--------|---|
| (3) Your gender (Mark with X in one block.) | Male | 8 |
| | Female | 9 |

| | | |
|---|-----------|----|
| (4) In which region did you complete Grade 12? (Mark with X in one block.) | Caprivi | 10 |
| | Erongo | 11 |
| | Hardap | 12 |
| | Karas | 13 |
| | Kavango | 14 |
| | Khomas | 15 |
| | Kunene | 16 |
| | Ohangwena | 17 |
| | Omaheke | 18 |
| | Omusati | 19 |
| | Oshana | 20 |
| | Otjikoto | 21 |
| Otjozondjupa | 22 | |
| Not in Namibia | 23 | |

| | |
|--|----|
| (5) What do you think a dictionary is? | 24 |
| | |
| | |
| | |

| | | |
|--|-----------------------|----|
| (6) How often did you use a dictionary at school? (Mark with X in one block.) | Almost every day | 25 |
| | Every week | 26 |
| | Once or twice a month | 27 |
| | Hardly ever | 28 |

| | | |
|---|---|----|
| Answer this question <i>only</i> if you answered "<i>Hardly ever</i>" to the previous question: (7) Why did you never really use a dictionary at school? (Mark with X in one block.) | There were no dictionaries available. | 29 |
| | We were not allowed to use dictionaries in class. | 30 |
| | There was one or only a few dictionaries in class, but I was afraid everyone would think I was stupid if I asked to use a dictionary. | 31 |
| | I do not know how to use a dictionary. | 32 |
| | I did not need or find it necessary to use a dictionary. | 33 |

| | | |
|--|--|----|
| (8) If you did use a dictionary at school, what kind of dictionary did you use? (Mark with X in one block.) | A monolingual dictionary, e.g. an English dictionary that explains the meanings of words in English by using definitions. | 34 |
| | A bilingual dictionary, e.g. an English dictionary that gives the translations of English words in another language, like Afrikaans, Damara>Nama or Oshiwambo. | 35 |
| | I don't know what kind of dictionary I used. | 36 |
| | I never really used a dictionary at school. | 37 |

| | |
|--|----|
| (9) If you did use a dictionary at school, what was the name of the dictionary that you used? If you can't remember, write "DON'T REMEMBER". | 38 |
| | |

| | | |
|---|-----|----|
| (10) Did any teacher ever teach you how to use a dictionary? (Mark with X in one block.) | Yes | 39 |
| | No | 40 |

| | | |
|--|-----------------------------|----|
| Answer this question <i>only</i> if you answered "Yes" to the previous question: (11) Which teacher taught you how to use a dictionary? (Mark with X in one block.) | One of my language teachers | 41 |
| | Another teacher | 42 |

| | | |
|--|-----|----|
| (12) Do you have your own dictionary? (Mark with X in one block.) | Yes | 43 |
| | No | 44 |

John is a first year student at UNAM. Below are a number of questions that he is wondering about. Which of the questions do you think a dictionary would answer for John? If you think that a dictionary would provide the answer, mark the "Yes" block next to the relevant question. If you think that a dictionary would not provide the answer to a question, mark the "No" block next to the relevant question. If you don't know whether a dictionary would provide the answer, mark the "Don't know" block next to the relevant question.

| | | |
|---|------------|----|
| (13) John is not sure how to spell the word <i>psychologist</i> . Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 45 |
| | No | 46 |
| | Don't know | 47 |

| | | |
|--|------------|----|
| (14) John wonders how the weather will be like tomorrow. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 48 |
| | No | 49 |
| | Don't know | 50 |

| | | |
|---|------------|----|
| (15) John wonders what the meaning of the word <i>courteous</i> is. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 51 |
| | No | 52 |
| | Don't know | 53 |

| | | |
|---|------------|----|
| (16) John must prepare for a class presentation and he is not sure how to pronounce the word <i>pterodactyl</i> . Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 54 |
| | No | 55 |
| | Don't know | 56 |

| | | |
|--|------------|----|
| (17) John wonders how much a hamburger costs. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 57 |
| | No | 58 |
| | Don't know | 59 |
| (18) John is writing and he is not sure where he can break off the word <i>interesting</i> at the end of a line. Is it <i>inte-resting</i> or <i>inter-esting</i> ? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 60 |
| | No | 61 |
| | Don't know | 62 |
| (19) John wonders what the plural form of the word <i>curriculum</i> is. Is it <i>curriculum</i> s or <i>curricula</i> ? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 63 |
| | No | 64 |
| | Don't know | 65 |
| (20) John reads the following sentence: "When Lisa's car broke down, she knew she was a sitting duck." John does not know the meaning of the expression <i>she was a sitting duck</i> . Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 66 |
| | No | 67 |
| | Don't know | 68 |
| (21) John is not sure what today's date is. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 69 |
| | No | 70 |
| | Don't know | 71 |
| (22) John is writing a birthday card for a friend. He is not sure whether he should write <i>Congratulations on your birthday!</i> or <i>Congratulations with your birthday!</i> Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 72 |
| | No | 73 |
| | Don't know | 74 |
| (23) John is writing a formal business letter to the Ministry of Education, and he is not sure whether he may use the word <i>stuff</i> in such a formal letter. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 75 |
| | No | 76 |
| | Don't know | 77 |
| (24) John is not sure whether he should spell the word <i>Christian</i> with a capital letter. Is it <u>C</u> hristian or <u>c</u> hristian? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 78 |
| | No | 79 |
| | Don't know | 80 |

| | | |
|---|------------|-----|
| (25) John wonders where the word <i>biltong</i> comes from. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 81 |
| | No | 82 |
| | Don't know | 83 |
| (26) John wants to know what the telephone number of the nearest hospital is. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 84 |
| | No | 85 |
| | Don't know | 86 |
| (27) John is not sure what the past tense form of <i>catch</i> is. Is it <i>catched</i> or <i>caught</i> ? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 87 |
| | No | 88 |
| | Don't know | 89 |
| (28) John wonders why Christmas Day always falls on 25 December. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 90 |
| | No | 91 |
| | Don't know | 92 |
| (29) John wonders what the abbreviation <i>VAT</i> stands for. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 93 |
| | No | 94 |
| | Don't know | 95 |
| (30) John wonders what sound a cat makes: Does a cat <i>bark</i> or does it <i>meow</i> ? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 96 |
| | No | 97 |
| | Don't know | 98 |
| (31) John wonders how many eggs are in a <i>dozen</i> : 10 eggs or 12 eggs? Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 99 |
| | No | 100 |
| | Don't know | 101 |
| (32) John wants to know the distance between Windhoek and Oshakati. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 102 |
| | No | 103 |
| | Don't know | 104 |

| | | |
|--|------------|-----|
| (33) John wonders whether the word <i>critic</i> is a verb or a noun. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 105 |
| | No | 106 |
| | Don't know | 107 |
| (34) John wonders what the opposite of the word <i>hot</i> is. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 108 |
| | No | 109 |
| | Don't know | 110 |
| (35) John is looking for a synonym for the word <i>nice</i> . Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 111 |
| | No | 112 |
| | Don't know | 113 |
| (36) John wants to find out how many books there are in the UNAM library. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 114 |
| | No | 115 |
| | Don't know | 116 |
| (37) John wonders if the word <i>bloody</i> is a swearword. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 117 |
| | No | 118 |
| | Don't know | 119 |
| (38) John wonders if he should put a comma before the word <i>because</i> in a sentence. Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 120 |
| | No | 121 |
| | Don't know | 122 |
| (39) John is not sure whether he should write <i>I am sick</i> or <i>I is sick</i> . Do you think a dictionary would provide the answer? (Mark with X in one block.) | Yes | 123 |
| | No | 124 |
| | Don't know | 125 |

Thank you very much for your cooperation!

Addendum B: Summary of the results of questionnaire items 13 to 39

| Item no. | Yes % | No % | Don't know % | n | X ² | df | p |
|----------|-------|------|--------------|-----|----------------|----|-------|
| 13 | 87.8 | 9.6 | 2.6 | 502 | 675.0 | 2 | <0.05 |
| 14 | 1.0 | 99.0 | 0.0 | 501 | 972.3 | 2 | <0.05 |
| 15 | 98.8 | 0.2 | 1.0 | 501 | 966.4 | 2 | <0.05 |
| 16 | 80.5 | 14.9 | 4.6 | 502 | 510.2 | 2 | <0.05 |
| 17 | 1.6 | 97.4 | 1.0 | 501 | 925.5 | 2 | <0.05 |
| 18 | 70.5 | 15.9 | 13.6 | 477 | 295.9 | 2 | <0.05 |
| 19 | 89.7 | 4.6 | 5.7 | 476 | 680.8 | 2 | <0.05 |
| 20 | 42.5 | 41.9 | 15.6 | 475 | 67.4 | 2 | <0.05 |
| 21 | 1.9 | 97.5 | 0.6 | 476 | 881.5 | 2 | <0.05 |
| 22 | 57.0 | 29.3 | 13.7 | 474 | 136.4 | 2 | <0.05 |
| 23 | 50.1 | 37.5 | 12.4 | 475 | 104.8 | 2 | <0.05 |
| 24 | 64.8 | 23.4 | 11.8 | 474 | 220.3 | 2 | <0.05 |
| 25 | 51.4 | 33.3 | 15.3 | 478 | 93.9 | 2 | <0.05 |
| 26 | 1.2 | 97.1 | 1.7 | 481 | 879.8 | 2 | <0.05 |
| 27 | 89.4 | 7.7 | 2.9 | 479 | 678.1 | 2 | <0.05 |
| 28 | 11.5 | 79.5 | 9.0 | 478 | 458.9 | 2 | <0.05 |
| 29 | 80.6 | 9.8 | 9.6 | 478 | 479.4 | 2 | <0.05 |
| 30 | 47.2 | 38.2 | 14.6 | 479 | 81.3 | 2 | <0.05 |
| 31 | 43.6 | 48.9 | 7.5 | 479 | 145.6 | 2 | <0.05 |
| 32 | 1.4 | 96.9 | 1.7 | 478 | 868.1 | 2 | <0.05 |
| 33 | 91.8 | 4.4 | 3.8 | 478 | 736.3 | 2 | <0.05 |
| 34 | 58.0 | 30.3 | 11.7 | 479 | 156.4 | 2 | <0.05 |
| 35 | 70.4 | 17.6 | 12.0 | 493 | 307.0 | 2 | <0.05 |
| 36 | 1.0 | 98.2 | 0.8 | 496 | 938.7 | 2 | <0.05 |
| 37 | 63.3 | 22.3 | 14.4 | 494 | 205.0 | 2 | <0.05 |
| 38 | 39.1 | 44.9 | 16.0 | 494 | 69.4 | 2 | <0.05 |
| 39 | 53.7 | 36.6 | 9.7 | 494 | 145.4 | 2 | <0.05 |

The Use of Pocket Electronic Dictionaries by Thai University Students

Atipat Boonmoh, *Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi, Bangmod, Thailand*
(*atipat.boon@kmutt.ac.th / atipat_b@yahoo.com*)

Abstract: This article reports on a small-scale study of Thai-speaking learners using pocket electronic dictionaries (PEDs) to read an English news article. It investigates how the subjects use their PEDs for reading comprehension. Thirty-nine undergraduate students completed a questionnaire survey. Of these, four were chosen to participate in the experiment. Observations and interviews were utilized to ascertain how the subjects used their PEDs. The findings showed superficial and partial reading of the dictionary entries. It also revealed several factors that may hinder dictionary look-up success. On the basis of this research, guidelines for buying PEDs and for teaching PED skills are proposed.

Keywords: DICTIONARY USE, POCKET ELECTRONIC DICTIONARY, DICTIONARY CONSULTATION, MONOLINGUAL DICTIONARY, BILINGUAL DICTIONARY

Opsomming: Die gebruik van sak- elektroniese woordeboeke deur Thaise universiteitstudente. Hierdie artikel doen verslag van 'n kleinskaalse studie van Thaisprekende aanleerders wat sak- elektroniese woordeboeke (SEW's) gebruik om 'n Engelse nuusartikel te lees. Dit ondersoek hoe die proefpersone hul SAW's gebruik vir leesbegrip. Nege-en-dertig voorgraadse studente het 'n vraelysopname voltooi. Uit hulle is vier gekies om aan die eksperiment deel te neem. Waarnemings en onderhoude is aangewend om te bepaal hoe die proefpersone hul SEW's gebruik het. Die bevindings het oppervlakkige en gedeeltelike lees van die woordeboekinskrywings getoon. Dit het ook verskeie faktore uitgewys wat woordeboekopsoeksukses mag belemmer. Op grond van hierdie navorsing word riglyne vir die koop van SEW's en vir die onderrig van SEW-vaardighede voorgestel.

Sleutelwoorde: WOORDEBOEKGEBRUIK, SAK- ELEKTRONIESE WOORDEBOEK, WOORDEBOEKRAADPLEGING, EENTALIGE WOORDEBOEK, TWEETALIGE WOORDEBOEK

1. Introduction

Educators usually recognize the reliability of the major printed English learner's dictionaries, but they are less likely to appreciate pocket electronic dictionaries (PEDs) since PEDs are normally developed and advertised in terms of their technology rather than their lexicographical features. PEDs are, how-

ever, common in South East Asian classrooms (Boonmoh and Nesi 2008, Deng 2005, Midlane 2005), their use being greatest in 'countries where PEDs are more cheaply and easily available ... and cultures in which microelectronic devices are extremely popular and fashionable' (Midlane 2005).

At King Mongkut's University of Technology Thonburi (KMUTT), Thailand, teachers have encouraged undergraduates to buy either the *Longman Active Study Dictionary* (LASD), or the *Cambridge Learner's Dictionary* (CLD), which is used as a key text in dictionary training skills courses. However, students seem to prefer to use PEDs when they are left to their own devices. A survey by Boonmoh and Nesi (2008) found that out of 1 211 students, 1 149 reported that they own at least one dictionary (95%). Of these, 938 (82%) own a monolingual dictionary in book form (presumably the LASD), 521 (45%) own a bilingual dictionary in book form, and 456 (40%) own PEDs. The survey indicated a mismatch, however, between the number of students who stated that they own monolingual print dictionaries (1 149) and the number who stated that they normally use them (46 for reading, and 103 for writing). The number of subjects who said that they own PEDs (456), on the other hand, was found to correspond quite well with the number of subjects who said that they normally use them (435 for reading, and 412 for writing). The students were required by their tutors to purchase the monolingual dictionary (the LASD), but PED use seems to have been the students' own choice.

1.1 Research in electronic dictionary use

Electronic dictionary use is an under-researched area (Jopling 2003, Tono 2000); and the few studies that look closely at what happens when electronic dictionaries are consulted mainly concern the use of learner's dictionaries on CD-ROM (e.g. Jopling 2003, Nesi and Hail 2002, Winkler 2001). Most prior research into PED use has been confined to quantitative ownership surveys (e.g. Deng 2005, Taylor and Chan 1994) and qualitative investigations into teacher and student attitudes and beliefs (e.g. Boonmoh and Nesi 2008, Sobkowiak 2002). A few simple experiments have been conducted, such as a look-up 'race' between PED and print dictionary users (Weschler and Pitts 2000), but there have been no close analyses of exactly what happens when PED users look up words. The 'Green tea' experiment is an attempt to fill in this research gap.

1.2 Research methods of researching PED use

Questionnaires and interviews have been among the most frequently used instruments for PED research (Boonmoh and Nesi 2008, Deng 2005, Sobkowiak 2002, Koren 1997, Tang 1997, and Taylor and Chan 1994) since they can be used as a way of obtaining results from a great number of respondents. They can be useful for identifying general trends which might then be examined more

closely in smaller, more empirical studies. A questionnaire alone, however, cannot reveal 'exactly what ... students are doing with their dictionaries, what they expect from them, and how easily they are satisfied during the process of consultation' (Atkins and Varantola 1998: 115). For this reason, it seems a good idea to triangulate questionnaire data with more qualitative data obtained by other means.

Observation is an obvious means of collecting data in educational settings. Observing PED use in a natural setting is almost impossible because PED consultation is a private activity, and one which learners are often inclined to be secretive about (Nesi and Haill 2002). Video recording users, a method employed by Jopling (2003) when investigating the use of CD-ROM dictionaries, may not be a practical means of researching PED use since PED users may often hunch over their PEDs when consulting them, and block the view of the video camera. Although 'spy' software has some potential as a means of observing online dictionary use (through keystroke logging and screenshots), it cannot be loaded into the standard PED. Although PED consultation is a private activity, observing its use may provide us with a better understanding of how students really use their PEDs.

2. Purposes of the study

The purposes of this study are to find out

- what type of dictionary students possess, use, and want to buy,
- how students use their PEDs to read an English news article, and
- how successful their PED consultations are.

3. Methodology

3.1 Subjects

The subjects of this study were 39 first-year undergraduate students who enrolled in a foundation English course (LNG 102 Fundamental English II) in the second semester of the academic year 2009 (November 2009–February 2010) at King Mongkut's University of Technology. They were from three departments: Computer Engineering, Mathematics and Multimedia. These subjects were chosen because they enrolled in the foundation English course which was taught by the researcher. Of 39 subjects, 4 were selected to take part in the experiment. These 4 subjects were chosen because they had claimed to possess and use pocket electronic dictionaries, and had indicated their willingness to participate in the experiment.

3.2 Research instruments

A questionnaire, interviews, and observations were employed in this study. Initially, a questionnaire entitled 'Dictionary Use of Thai University Students' was completed by 39 subjects. The questionnaire consisted of 3 questions whose aim was to find out students' dictionary ownership, preference for dictionary use, and preference for dictionary purchase. Based on the findings of the questionnaire, only those who reported that they possess PEDs (10 subjects) were asked to participate in the study. Only 4 subjects expressed an interest in the study. Therefore each of them, in an individual session, was asked to complete a reading task. Subjects were asked to read a news article and orally reported what the article was about. This article was taken and adapted from a BBC website <http://news.bbc.co.uk/2/hi/health/8453628.stm> entitled 'Green tea "may block lung cancer"' which discusses that drinking green tea may avoid the risk of lung cancer.

This text had been piloted with a comparable group of volunteers, and had proved to be appropriate in terms of topic, difficulty level and length.

Subjects were asked to read the article in individual sessions, in the presence of the researcher who also observed every word looked up and completed an observation check sheet. The check sheets were used to help him record the words and meanings the subjects looked up, and were also used as notes to ask some specific questions during the retrospective interviews conducted with each subject at the end of the session.

Green tea 'may block lung cancer'

Drinking green tea may offer some protection against lung cancer, say experts who studied the disease at a medical university in Taiwan.

The latest work in more than 500 people adds to growing evidence suggesting the beverage has anti-cancer powers. In the study, smokers and non-smokers who drank at least a cup a day cut their lung cancer risk significantly, a US cancer research conference heard. The protection was greatest for people carrying certain genes. But cancer experts said the findings did not change the fact that smoking is bad for health.

Green tea is made from the dried leaves of the Asian plant *Camellia sinensis* and is drunk widely across Asia. The rates of many cancers are much lower in Asia than other parts of the world, which has led some to link the two. Laboratory studies have shown that extracts from green tea, called polyphenols, can stop cancer cells from growing. But results from human studies have been mixed. Some have shown a protective effect while others have failed to find any evidence of protection.

In July 2009, the Oxford-based research group Cochrane published a review of 51 studies on green tea and cancer which included over 1.5 million people. They concluded that while green tea is safe to drink in moderation, the research so far is conflicting about whether or not it can prevent certain cancers. (229 words)

3.3 Data analysis

Findings were derived from consideration of three data sources: the questionnaire, observation notes, and interviews. The data was analyzed only when the subjects consulted their PEDs to comprehend the English passage. The number of words looked up, and the number of successful and unsuccessful look-ups were noted. The data from the interviews, and the observation check sheets were analyzed with reference to the following questions:

- How did the subjects write their summaries?
- What words did they look up?
- Did they search for words in the appropriate word class?
- Did they find the words they were looking for?

4. Result

4.1 Student's use of dictionaries

Table 1: Dictionary Use of Thai University Students

| | |
|--|--|
| Gender | 17 Male 22 Female |
| Department | 7 Mathematics 26 Engineering 6 Multimedia |
| 1. Dictionaries you own | 32 Monolingual dictionary in book form 35 Bilingual dictionary in book form 10 Pocket electronic dictionary 9 Monolingual dictionary in CD form |
| 2. Dictionaries you normally use | 4 Monolingual dictionary in book form 30 Bilingual dictionary in book form 9 Pocket electronic dictionary 4 Monolingual dictionary in CD form |
| 3. Type of dictionaries you plan/would like to buy | 22 Pocket electronic dictionary 13 Bilingual dictionary in book form 4 Monolingual dictionary in book form |

It can be seen from Table 1 that of 39 respondents, 35 (90%) owned a bilingual dictionary in book form, 32 (82%) owned a monolingual dictionary in book form which was a recommended purchase by the School (78% wrote 'Cambridge', probably referring to CLD), and 10 (26%) owned PEDs. The CLD on CD-ROMs are 'bundled' with the printed copies of CLD on sale at KMUTT.

However, only 9 respondents claimed ownership of a monolingual dictionary on CD-ROM. This may suggest that many respondents had not explored the inside of their CLD copies.

For the second question, there was a 90% drop in the number of respondents who used a monolingual dictionary. Although 32 respondents claimed to own a monolingual dictionary, only 4 respondents (10%) reported that they normally use it. This may be taken to imply that the ownership of a certain type of dictionary does not equate to its use. However, the number of respondents who reported owning a bilingual dictionary in book form and a PED seem to correspond with the number of respondents who reported using it.

The last question enquired about the respondents' preferences when buying a new dictionary. More than half of the participants, 56% (22 out of 39), reported that they would like to buy PEDs, 13 respondents wanted to buy a bilingual dictionary in book form, and only 4 students (12.5%) wanted to buy a monolingual dictionary in book form. None stated that they would want to buy dictionaries on CD-ROM.

The findings from this survey provided a general picture of how participants used their dictionaries. The findings inform us about the dictionaries the participants owned, which dictionaries they preferred to use or normally used, and which dictionaries they would like to buy in future.

4.2 Students' reading behaviours and look-up analysis

The ways these 4 subjects read the article were obvious. All of them started looking up unknown words from the beginning. None of them read a second time. Subject A, B and C read the article sentence by sentence and looked up unknown words while Subject D read the article word by word and tried to understand the word before moving to another word.

A PED normally contains at least three dictionaries, i.e. a Thai–English, an English–Thai, and an English–English dictionary. The sources of the three dictionaries differ among PED brands (e.g. *CyberDict* vs. *TalkingDict*) and PED models (e.g. *CyberDict 3* vs. *CyberDict 4*). In this study, Subject A, B, and C used PEDs by *CyberDict* and Subject D used PEDs by *TalkingDict*. It was found that all 4 subjects consulted only the English–Thai PED dictionary. This was probably because the English–Thai dictionary was set up as the default when a PED is turned on. The interview also suggested that the subjects were not familiar with the English–English dictionary and they thought that looking up words from an English–English dictionary was time-consuming. Some stated that they could not understand the English definitions and they had to use the English–Thai dictionary in order to understand some English words in the definitions.

As can be seen in Table 2, the numbers of dictionary look-ups per subject ranged from 22 (Subject C) to 40 (Subject D). Subject D looked up one in every five words (40 look-ups out of 229 words) in the English news article. This might have reflected the fact that Subject D tackled reading the article word for word.

Table 2: Subjects' look-up analysis

| Subject | Number of look-ups | Words placed in the wrong order | Inappropriate words searched | Unsuccessful look-ups |
|-----------------------------|--------------------|-----------------------------------|---|---|
| A (<i>CyberDict</i>) | 28 | research | leaves, suggesting, conflicting, published | – |
| B (<i>CyberDict</i>) | 30 | research, may, rate | leaves, expects, extracts, published, adds, conflicting | – |
| C (<i>CyberDict</i>) | 22 | research, review, | extracts, published, suggesting | – |
| D (<i>TalkingDict</i>) | 40 | link, research, rate, may, review | leaves, smokers, experts, extracts, adds, suggesting, conflicting, growing, published, included | smokers, experts, extracts, adds, suggesting, published |
| Total | 120 | 11 | 23 | 6 |

Out of 120 look-ups, 11 (9.1%) were misidentified for their word class. The subjects might not have been aware that one word form could have more than one word class. For example, 'research' can function as a verb and also a noun. The subjects might not have analyzed the part of speech where the word 'research' occurred in context. A second reason for this unidentified word class would have been because of the PEDs the subjects used.

Locating inappropriate words to look up is a major problem for dictionary consultation. It is seen that out of 120, 23 words (19.2%) were inappropriately chosen to look up. It seems that the subjects copied the exact word form from the article to look up in their PEDs, without analyzing the part of speech or adjusting to remove inflections. For example, all 4 subjects looked up the word 'published' ('the Oxford-based research group Cochrane published a review of 51 studies') without adjusting this word to its root form ('publish'). Subject D selected the most number of inappropriate words to look up. These include: 'leaves', 'smokers', 'experts', 'extracts', 'adds', 'suggesting', 'conflicting', 'growing', 'published', 'included'.

There are 6 look-ups that were unsuccessful. This was the result of failure in adjusting inflected forms appropriately. It should be noted that all of the unsuccessful look-ups were looked up by Subject D. This may at first be taken to imply that Subject D was the least skilful in PED use. However, after a careful investigation of the PED, it was found that the cause of these unsuccessful look-ups could also have been the result of the PED brand the subject used. The investigation of the two PED brands (*TalkingDict* and *CyberDict*) revealed two important aspects: the content and the function.

It is interesting to note that the words 'conflicting' and 'growing' are listed as headwords in the *TalkingDict* (as adjectives). However, these two words did not exist as headwords in major English–English dictionaries (e.g. *Longman Dictionary of Contemporary English on CD-ROM* and *Collins Cobuild Advanced Learner's Dictionary on CD-ROM*).

The content of the English–Thai dictionary in *CyberDict* comes from Thiengburanathum's (1998) *SE-ED's Modern English–Thai Dictionary (Complete and Updated) Desk Reference Edition*, Bangkok: SE-Education. On the other hand, the content of the English–Thai dictionary in *TalkingDict* is taken from the *English–Thai Dictionary* (compiled by lecturers from the Chalermprakit Centre of Translation and Interpretation). Note that no authors' titles or years of publication were provided.

CyberDict has a unique function that can display the most appropriate root form. For example, when typing 'suggests', 'suggesting', or 'suggested', the PED displays 'suggest'. This function, however, is not present in *TalkingDict*. Had Subject D used *CyberDict* to look up these six words, his searches would have been successful.

5. Conclusion and Discussion

The findings from the questionnaire survey suggest a growing trend for pocket electronic dictionaries. The findings accord with the findings of previous studies that PED use will increase (Taylor and Chan 1994, Midlane 2005). Only a few participants would like to buy monolingual dictionaries. This may be taken to imply that the use of monolingual dictionaries was imposed on the participants and if they had their own choice to choose what dictionaries to buy, they might not have bought monolingual dictionaries as recommended by the Department.

The findings from the 'Green tea' experiment revealed insights into how students use PEDs. The problems the subjects encountered revealed a lack of reading skills and dictionary-using skills. The subjects lacked strategies in tackling the reading task. Some did not perform any reading strategies, e.g. looking at the title, skimming, or reading the main ideas, guessing words from context. As a consequence, the lack of reading strategies had led the subjects to rely heavily on searching for equivalent meanings from the PEDs. The subjects' look-up behaviours, however, were found to be partial. Some subjects were not

aware of the class word and inflected form of the words. Some of them did not select appropriate word forms to look up. However, the reason the subjects did not select the appropriate forms might have been the result of the difference between L1 (i.e. Thai) and L2 (i.e. English). For example, Thai does not have inflected forms (e.g. manage, manages, managed, managing) as in English. Thai derivational morphemes are prefixes rather than suffixes, e.g. **จัดการ** [chatkan] (verb 'manage'), **การจัดการ** [kanchatkan] (noun 'management') **ผู้จัดการ** [phu-chatkan] (noun 'manager'). The subjects might not have been aware of this difference and this could affect dictionary use. Moreover, the PED brand was found to be an important factor hindering or promoting look-up success.

6. Recommendation

Since PEDs will continue appearing in the classroom, teachers should provide authoritative advice on how to buy PEDs and/or train students to use PEDs. Based on the findings of this research, the following guidelines for buying PEDs and for teaching PED skills are proposed: Teachers may ask students to explore and compare the dictionary contents of the PED the students would like to buy in terms of lexicographical features (e.g. name of dictionaries, their authors' titles, years of publication, and the edition of the specific dictionary), and technological features (e.g. an ability to display the root form, wild card search). Teachers may teach students to make sure that they know how to identify the part of speech of words, how to adjust inflected forms to find root forms, and how to analyze word classes. When teaching English to a multilingual class, it is useful for teachers to teach fundamental facts about the L1 and the L2 (i.e. English), so that they know how differences can affect the way they will use their PEDs.

References

- Atkins, B.T.S. and K. Varantola. 1998. Monitoring Dictionary Use. Atkins, B.T.S. (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 83-122. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Boonmoh, A. and H. Nesi. 2008. A Survey of Dictionary Use by Thai University Staff and Students, with Special Reference to Pocket Electronic Dictionaries. *Horizontes de Lingüística Aplicada* 6(2): 79-90.
- Deng, Y.P. 2005. A Survey of College Students' Skills and Strategies of Dictionary Use in English Learning. *CELEA Journal* 28(4): 73-77.
- Jopling, F. 2003. *Towards a New Methodology for Research on Electronic and Paper-Based Dictionary Use: A Pilot Study of Look-Up Patterns on Screen and in Print*. Unpublished M.A. Thesis. Coventry: University of Warwick.
- Koren, S. 1997. Quality versus Convenience: Comparison of Modern Dictionaries from the Researcher's, Teacher's and Learner's Points of View [online]. *TESL-EJ* 2(3), January.

- Midlane, V.** 2005. *Students' Use of Portable Electronic Dictionaries in the EFL/ESL Classroom: A Survey of Teacher Attitudes*. M.Ed. Dissertation. Manchester: University of Manchester, Faculty of Education.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
- Sobkowiak, W.** 2002. The Challenge of Electronic Learner's Dictionaries [online]. *Teaching English with Technology: A Journal for Teachers of English* 2(1), January 2007. http://www.iatefl.org.pl/call/j_article7.htm#sob.
- Tang, G.** 1997. Pocket Electronic Dictionaries for Second Language Learning: Help or Hindrance? *TESL Canada Journal* 15(1): 39-57.
- Taylor, A. and A. Chan.** 1994. Pocket Electronic Dictionaries and their Use. Martin, W., W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg and P. Vossen (Eds.). 1994. *Euralex '94 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 598-605. Amsterdam: Free University Press.
- Tono, Y.** 2000. On the Effects of Different Types of Electronic Dictionary Interfaces on L2 Learners' Reference Behaviour in Productive/Receptive Tasks. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*: 855-862. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Stuttgart University.
- Weschler, R. and C. Pitts.** 2000. An Experiment Using Electronic Dictionaries with EFL Students [online]. *The Internet TESL Journal* 6(8). <http://iteslj.org/Articles/Weschler-ElectroDict.html>.
- Winkler, B.** 2001. English Learners' Dictionaries on CD-ROM as Reference and Language Learning Tools. *ReCall* 13(2): 191-205.

Appendix: Dictionary Use of Thai University Students

Please answer this questionnaire to find out more about your dictionary using habits. The information given will be kept confidential and answering this questionnaire will NOT affect your grades.

Please answer the following questions by ticking (✓) the appropriate spaces provided or answering in the spaces given.

(Personal information)

Faculty of study:

Major of study:

Email address:

Do you own/have any dictionaries?

Yes (Please continue) No (Please answer question 3)

(1) Which of these dictionaries do you have? **You may tick more than one item.**

(a) Pocket electronic dictionary (e.g. *TalkingDict*, *CyberDict*, *VTech*)

Please specify which brand you have

(b) Bilingual dictionary in book form

(c) Monolingual dictionary in book form

Please specify which brand you have

(d) Monolingual dictionary in CD form

(e) Other(s)

Please specify

(2) Which dictionary do you prefer to use? Refer to the type of dictionary in question 1.

Type of dictionary

(3) If you do not have a dictionary or if you plan to buy a new dictionary, what type of dictionary would you like to buy? Refer to types of dictionaries in question 1.

Type of dictionary

Thank you for you cooperation.

Borrowing and Dictionary Compilation: The Case of the Indigenous South African Languages*

Munzhedzi James Mafela, *Department of African Languages, University of
South Africa, Pretoria, South Africa (mafelmj@unisa.ac.za)*

Abstract: Borrowing occurs when a language adds a word from another language to its own lexicon. Languages in neighbouring regions often borrow from each other. In South Africa, there are over nine indigenous languages, in addition to Afrikaans and English, all of which coexist. In their coexistence they borrow from one another. African language dictionaries reveal that these languages have borrowed a great deal from Afrikaans and English. However, one would expect these dictionaries to reflect some borrowing from other indigenous languages; on the contrary, they include very few such words. Although, one encounters many words from indigenous languages being used by other African languages in their spoken language and literary works, these are not included in dictionaries. This article seeks to highlight factors contributing to the failure to include words borrowed from other indigenous languages in African language dictionaries.

Keywords: BORROWING, DICTIONARY COMPILATION, INDIGENOUS LANGUAGES, LEXICON, MORPHEME, VOCABULARY, DEVELOPING LANGUAGES, LOAN WORDS, TERMINOLOGY, ETYMOLOGY, LEXICOGRAPHY

Opsomming: Ontlening en woordeboeksamestelling: Die geval van inheemse Suid-Afrikaanse tale. Ontlening vind plaas wanneer 'n taal 'n woord uit 'n ander taal by sy eie leksikon voeg. Tale in aangrensende streke leen dikwels van mekaar. Daar is meer as nege inheemse tale in Suid-Afrika, bo en behalwe Afrikaans en Engels, wat saam bestaan. In hulle naasbestaan leen hulle van mekaar. Afrikataal-woordeboeke getuig daarvan dat hierdie tale heelwat uit Afrikaans en Engels geleen het. Daar sou egter verwag kon word dat hierdie woordeboeke sal getuig van ontlening uit ander inheemse tale; in teendeel, hulle sluit baie min van sulke woorde in. Alhoewel baie woorde van inheemse tale in ander Afrikatale gebruik word in die gesproke taal en letterkundige werke, is hierdie woorde nie in die woordeboeke opgeneem nie. Die doel van hierdie artikel is om die faktore uit te lig wat bydra tot die versuim om leenwoorde van ander inheemse tale in die Afrikataal-woordeboeke op te neem.

* This article was presented as a paper at the Fourteenth International Conference of the African Association for Lexicography, organised by the Xhosa Department, University of the Western Cape, Bellville, South Africa, 6–8 July 2009.

Sleutelwoorde: ONTLENING, WOORDEBOEKSAMESTELLING, INHEEMSE TALE, LEKSIKON, MORFEEM, WOORDESKAT, ONTWIKKELENDE TALE, LEENWOORDE, TERMINOLOGIE, ETIMOLOGIE, LEKSIKOGRAFIE

1. Introduction

Languages are not static but develop over time. During this development, words may be added or lost: while new vocabulary is accepted into a language, some words may fall into disuse and become obsolete. This may be the result of contact between language groups and the introduction of new styles of living in the society. As Fromkin and Rodman (1998: 459) note: "Changes in the lexicon also occur, including the addition of new words, changes in the meanings of words, and loss of words." Words may enter a language in many ways. Fromkin and Rodman (1998) mention compounding, derivational processes, coinage, recombining old words to form new ones with new meanings, and borrowing.

The discussion in this article focuses on the concept of *borrowing*. The term *borrow* has various meanings. It can be associated with taking something on the understanding that it will be returned, or replaced with something similar. It may also mean to adopt something as one's own, or to adopt a word from another language (Guralnik 1981). In this study, the word is used in the sense of the third definition, i.e. to adopt a word from another language. Borrowing, in this regard, occurs when a language adds a word or morpheme from another language to its own lexicon. Fromkin and Rodman (1998: 459) describe it as follows:

Borrowing words from other languages is an ... important source of new words. Borrowing occurs when one language adds to its own lexicon a word or morpheme from another language, often altering its pronunciation to fit the phonology rules of the borrowing language.

Mafela (1996: 165) mentions that

Adoption is brought about by the contact between people who speak different languages. One language tends to adopt words from the other.

Borrowing can therefore be defined as a word or phrase in one language adopted by another. Haugen (1950: 213-215) defines a loan word as follows:

A loan word is a complex form composed entirely of foreign morphemes. Speakers importing a loan word import not only the meaning of the word but also its phonemic shape although the substitution of foreign formative phonemes may be more or less complete.

Indigenous languages in South Africa have been enriched through this process of borrowing. With the introduction of terms for new political, economic, social

and religious activities, the vocabulary of the African languages has been significantly enhanced. Many new words have been adopted from Afrikaans and English. Although the indigenous peoples of South Africa share a great deal regarding culture, their languages are distinct from one another. However, owing to social conditions in the country, language communities come into contact in several environments such as places of work. During this interaction, they borrow vocabulary from one another. Nevertheless, the borrowed vocabulary is rarely reflected in indigenous African language dictionaries. Afrikaans and English dictionaries are more comprehensive in this regard as they include words adopted from the African languages, for example *mopanie(boom)* (Afrikaans), *mopane/mopani* (English), i.e. *mupani* in Tshivenda, *lobôla* (Afrikaans), *lobola* (English), i.e. *lobola* in isiZulu, and *maroela(boom)* (Afrikaans), *marula* (English), i.e. *morula* in Sesotho sa Leboa, to mention only a few. This article seeks to highlight factors contributing to the failure to reflect borrowings in indigenous African language dictionaries, and the importance of the inclusion of such loan words in these dictionaries.

2. Borrowing from developed languages

In this article, the indigenous languages referred to are the neighbouring languages isiNdebele, isiXhosa, isiZulu, Sesotho, Sesotho sa Leboa, Setswana, Siswati, Tshivenda and Xitsonga. Neighbouring languages will always borrow from each other. No two languages come into contact without influencing each other. An extensive number of words can be traced to another language. In many instances, the developing language will borrow more words from the developed. As the indigenous African languages are developing languages in many terminological fields, they are therefore borrowers. These languages have borrowed much of their terminological vocabulary from Afrikaans and English, owing to the fact that Afrikaans and English are more developed in this respect. Because the indigenous African languages coexist with Afrikaans and English, much scientific, economic, political and religious terminology has been borrowed by them. If one takes a look through some of the dictionaries covering the indigenous African languages, one notices that the words listed below, for instance, have been accepted as part of the lexicon of these languages. This list contains some of the words giving evidence as having been borrowed by and adapted to the indigenous African languages:

Tshivenda:

| | |
|------------------------------|------------------|
| iron (English) | <i>aini</i> |
| altar (English) | <i>alitare</i> |
| bank (Afrikaans and English) | <i>bannga</i> |
| brood (Afrikaans) | <i>vhurotho</i> |
| papier (Afrikaans) | <i>bammbiri</i> |
| deposit (English) | <i>diphositi</i> |

| | |
|-----------------------|--------------------|
| bicycle (English) | <i>baisigira</i> |
| mathematics (English) | <i>mathematiki</i> |
| wine (English) | <i>waini</i> |
| kilometre (English) | <i>khilomitha</i> |

Nguni languages:

| | |
|------------------------------|-----------------------------|
| petrol (English) | <i>ipetroli</i> (isiXhosa) |
| botter (Afrikaans) | <i>ibhotolo</i> (isiXhosa) |
| bank (Afrikaans and English) | <i>ibhanki</i> (isiXhosa) |
| paper (English) | <i>iphepha</i> (isiXhosa) |
| deposit (English) | <i>dipozitha</i> (isiXhosa) |
| library (English) | <i>ilabhulali</i> (isiZulu) |
| zoology (English) | <i>izoloji</i> (isiXhosa) |
| dorp (Afrikaans) | <i>idolobha</i> (isiZulu) |
| science (English) | <i>isayensi</i> (isiZulu) |
| percent (English) | <i>iphesenti</i> (isiZulu) |

Sotho languages:

| | |
|------------------------------|--|
| iron (English) | <i>aene</i> |
| bank (Afrikaans and English) | <i>banka</i> (Sesotho), <i>panka</i> (Sesotho sa Leboa) |
| papier (Afrikaans) | <i>pampiri</i> |
| petrol (English) | <i>peterole</i> (Sesotho sa Leboa) |
| mathematics (English) | <i>mathematiki</i> (Sesotho sa Leboa) |
| tafel (Afrikaans) | <i>tafole</i> (Setswana), <i>tafola</i> (Sesotho sa Leboa) |

Xitsonga:

| | |
|------------------------------|-------------------|
| iron (English) | <i>ayini</i> |
| bank (Afrikaans and English) | <i>banki</i> |
| paper (English) | <i>phepha</i> |
| kilometre (English) | <i>khilomitha</i> |
| science (English) | <i>sayense</i> |
| dorp (Afrikaans) | <i>doroba</i> |

3. Borrowing from indigenous languages

Because of the coexistence of indigenous languages in South Africa, one would expect that dictionaries of these languages would reflect many loan words from one another. Indigenous languages of South Africa share a great deal of vocabulary, as well as some cultural aspects, because they belong to the same language family. For example, the word for *person* in all South Africa's indigenous languages is similar, namely *muthu* (Tshivenda), *motho* (Sotho languages), *munhu* (Xitsonga), *umntu* (isiXhosa), *umuntu* (isiZulu) (Reynierse 1991: 107). The same holds for the term *woman*. Its equivalents in the different languages are as follows: *musadzi* (Tshivenda), *mosadi* (Sotho languages), *nsati* (Xitsonga),

umfazi (isiXhosa and isiZulu). However, these languages are also individual languages, differing in many cultural aspects. These differences between language groups are often influenced by their geographical location. Each indigenous language in South Africa is concentrated in a particular geographical area characterised by particular geographical conditions. Differences in culture and geographical location lead to differences in the use of certain terminology.

Indigenous languages borrow terminology from each other in their spoken language and their literary texts. But while it is common to encounter words borrowed from another indigenous language, it is rare to find them included in the dictionary of that particular indigenous language. This suggests that there is a problem, and this problem seems to lie with the lexicographers.

An example is the word *mukhukhu*, which appears to have originated among the Sotho people, specifically the Basotho ba Leboa. It refers to a building made of corrugated iron, or a temporary shelter. The door is so low that when people want to enter, they have to crouch (*khukhuna*). This structure occurs mostly in urban areas where people of all indigenous language groups coexist. This word has been adopted by all South African indigenous languages. As it is in popular usage, one would expect to find *mukhukhu* in all African language dictionaries. However, it does not form part of the lexicon in any of the checked Tshivenda dictionaries. Instead, the word *mushasha* has been included. Although, according to Van Warmelo (1989), both *mukhukhu* and *mushasha* have the same meaning, *mushasha* is a temporary shelter of branches and grass, meant for travellers, whereas *mukhukhu* is a temporary shelter made of corrugated iron. This suggests that *mukhukhu* should be part of the words in an indigenous African language dictionary. In addition, the meaning of *mukhukhu* can also be linked to a type of dance practised by almost all indigenous language speakers in this country. This dance is performed by members of the Zion Christian Church, a church not restricted to a particular ethnic group, but one which embraces all South Africans as well as foreigners from neighbouring countries. This word has been accepted into the spoken indigenous languages in South Africa. It is therefore surprising to find that it is not included in the dictionaries of some of these languages.

Another example is the word *shonzha*, peculiar to Vhavenda. It is defined by Van Warmelo (1989: 336) as follows:

shonzha 5 sp. A large edible caterpillar (of the moth *Nudaurelia belina* Westw.), which in season is collected in great numbers from trees in the dry north country, dried, and eaten as a rare and expensive delicacy

These caterpillars, *mashonzha*, collected from the *mupani* tree found in the northern part of Venda, are regarded as a delicacy by the Vhavenda. The word *shonzha* is therefore part of the Tshivenda lexicon. But the Vhavenda are not the only people who enjoy eating *mashonzha*; neighbouring ethnic groups such as the Vatsonga and the Basotho ba Leboa do too. It is therefore clear that the term *shonzha* should be included in dictionaries of these indigenous languages.

However, except in Tshivenda, no entry was found in any dictionary of the South African indigenous languages. Today people from all backgrounds come into contact in urban areas because of the movement of people all over the country. Many language groups are thus familiar with *mashonzha*. The word *shonzha* should therefore be included in dictionaries as it has been accepted in various spoken languages.

A further example of a word used in several languages is *futhi*, a Nguni word which has been adopted by Tshivenda. Doke, Malcolm and Sikakana (1958: 67) define it as follows:

futhi (adv.) 1. Again, once more. 2. In addition, also, too.

In their daily conversation, Vhavenda use *futhi* to denote "again". In standard Tshivenda "again" is translated by the word *hafhu*. However, many Vhavenda use the adopted *futhi* instead of *hafhu*. This is illustrated in a text taken from *Musandiwa na khotsi Vho-Liwalaga* (Maumela 1973: 37), which reads:

Tshi re hone ene o tou nala, a sokou ŋuwa, futhi na u onesa a songo nnyonesa nŋe khotsi awe.

(She left on her own, in addition she did not even bid me, her own father, a goodbye.)

In this passage, *futhi* has been used to mean "in addition". The mere fact that the author has used this word in a literary text, and that the word has been accepted by the evaluators of the book, means that it has been adopted into spoken Tshivenda and is frequently used. Although Tshivenda has an equivalent for *futhi*, this does not mean that it should not be reflected in Tshivenda dictionaries.

Because of its geographical location, Tshivenda shares much vocabulary with Xitsonga and Sesotho sa Leboa. While Xitsonga is spoken in the eastern part of Venda, Sesotho sa Leboa is used in the western and southern parts. In addition, these ethnic groups come into contact in the workplace and in their social lives. The vocabulary they share is realised mainly in literary texts and spoken language. Words such as *vhuthada* and *mutsheka* are good examples in this regard.

Vhuthada seems to have been borrowed from the Sotho languages (*bothata*) (Reynierse 1991: 73), most particularly Sesotho sa Leboa owing to its proximity. In his literary text, Maumela (1973: 35) uses it as follows:

Mafhungo aya a ŋu khezwi vhu vhuthada. Ndi hune a nga tou itwa hani na?

(This matter of yours seems to be difficult. How is it going to be solved?)

Although the Tshivenda equivalent is *vhuleme*, the use of *vhuthada* indicates that it has been accepted by the community. However, *vhuthada* is not found in Tshivenda dictionaries in spite of this acceptance. Its use in literary texts is a good reason for its inclusion in Tshivenda dictionaries.

A similar example is the Xitsonga word *nceka* which is in popular use among the Vhavenda. A *nceka* is an item of clothing worn by both Vatsonga men and women. The word is defined as follows (Cuenod 1967: 119):

nceka 3, piece of material worn tied round the waist by men, round the shoulders by women.

When the Vhavenda refer to this type of clothing, they call it *mitsheka*. Some lexicographers may argue that the equivalent of *nceka* is *niwenda* in Tshivenda. A *niwenda* is a special type of garment made for Vhavenda women, but it is not similar to a *mitsheka* although they both have the same function. The word *niwenda* is defined by Van Warmelo (1989) as follows:

niwenda 3 (pl. miñw) female upper garment of salem-pore, just a length of cloth with strip (bannda) sewn on crosswise at the top to make it longer, and with two tapes (mivhofho) of the same material to tie over the shoulder.

As in the case of *vhuthada*, *mitsheka* has become accepted in Tshivenda literary texts. Maumela (1974: 69) writes as follows:

Ngeno fhasi mikumba i tshi qi vha hone yo ingwa ntha ha mitsheka.
(At the waist, he puts on skins on top of the salem-pores.)

The person referred to in this text is a male diviner. He wears skins and salem-pores (*mitsheka*) around his waist. He could not have put a *niwenda* around his waist because it is not meant to be worn by men. This would have caused the readers to form a wrong impression. As indicated in Van Warmelo's definition, the *niwenda* is used to cover the whole body, from the shoulders down, whereas the *mitsheka* is mostly used either as a skirt (to cover the body from the waist downward) or as a shawl around the shoulders. It is therefore necessary to include the word *mitsheka* in Tshivenda dictionaries, because it has been accepted in both the spoken and the written language.

4. Factors contributing to the non-borrowing by one indigenous African language from another

Lexicographers of the South African indigenous languages are reluctant to include loan words from other indigenous languages in their dictionaries for various reasons. Firstly, they may be hesitant to find the etymological meanings of words adopted from these languages. Borrowing is closely connected with etymology. Where the meaning of a borrowed word is not clear, lexicographers must trace its origin to the source language. Etymology will provide a clearer understanding of what the word means. Lexicographers should be prepared to undertake this type of research, drawing upon the cultures of other indigenous African languages, if they are really intent on enriching the language with which they are dealing.

Secondly, lexicographers are unwilling to include loan words from other indigenous languages because they want to maintain the purity of their own languages. Thinking that they will corrupt their own languages, they try to preserve the standard language. On the other hand, these lexicographers seem unconcerned about including words borrowed from Afrikaans and English. The idea of corrupting a language is only considered when it comes to borrowing from fellow indigenous African languages. This attitude must be changed if the indigenous African languages are to be developed and promoted. Lexicographers should recognise that language changes in course of time as Fromkin and Rodman (1998: 409) stress:

No academy and no guardians of language purity can stem language change, nor should anyone attempt to do so since such change does not mean corruption.

Lastly, lexicographers do not include loan words from other indigenous African languages in their dictionaries because there is an element of disparagement towards these languages. They are therefore hesitant to include vocabulary from these languages in their dictionaries, notwithstanding that particular words have been accepted into the spoken language and literary texts of their own languages. Lexicographers prefer to include loan words from well-developed languages such as Afrikaans and English in their dictionaries. Western civilization has introduced foreign lifestyles which are accompanied by new vocabulary which should be included in indigenous African language dictionaries. However, it is also important to include vocabulary from other indigenous South African languages when it has become accepted in the spoken and written language. The inclusion of these loan words will assist dictionary users in learning more about the culture of other indigenous African languages.

5. Conclusion

Borrowing vocabulary from one language by another is crucial as languages all over the world are enriched by this process. However, this should not be restricted to borrowing from developed languages only. In South Africa, borrowing among the indigenous languages should also be encouraged.

The inclusion of lexical items from one indigenous language in another will enrich the languages in many ways. The vocabulary of these languages will expand, facilitating communication between the different language groups because they will share words. If words are borrowed from one indigenous language by another and defined properly, dictionary users will be exposed to other cultures which will further enhance communication and act as a unifying element among cultural groups. Developing and promoting indigenous languages should not be measured only by how much has been adopted from developed languages such as Afrikaans and English, but should consider also the terminology taken from other indigenous languages. The time has come for

lexicographers of indigenous languages to start compiling bilingual and trilingual dictionaries of the South African indigenous languages.

References

Dictionaries

- Cuenod, R.** 1967. *Tsonga-English Dictionary*. Braamfontein: Sasavona Publishers and Booksellers.
- Doke, C.M., D. McK. Malcolm and J.M.A. Sikakana.** 1958. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- Guralnik, D.B. (Ed.)**. 1981. *Webster's New World Dictionary of the American Language*. Student Edition. Engelwood Cliffs, N.J.: Simon and Schuster.
- Reynierse, C. (Ed.)**. 1991. *South African Multi-Language Dictionary and Phrase Book*. Cape Town: The Reader's Digest Association South Africa.
- Van Warmelo, N.J.** 1989. *Venda Dictionary: Tshivenda-English*. Pretoria: J.L. van Schaik.

Other sources

- Fromkin, V. and R. Rodman.** 1998. *An Introduction to Language*. Sixth Edition. Orlando: Harcourt Brace College Publishers.
- Haugen, E.** 1950. The Analysis of Linguistic Borrowing. *Language* 26: 210-231.
- Mafela, M.J.** 1996. *The Elements of Fiction in the Novels of T.N. Maumela*. Pretoria: Kagiso Publishers.
- Maumela, T.N.** 1974. *Vhavenḁa Vho-Matshioha*. Johannesburg: APB.
- Maumela, T.N.** 1968. *Musandiwa na Khotsi Vho-Liwalaga*. Pretoria: J.L. van Schaik.

Borrowing and Loan Words: The Lemmatizing of Newly Acquired Lexical Items in Sesotho sa Leboa*

V.M. Mojela, *Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, South Africa (mojelav@ul.ac.za)*

Abstract: The influence of foreign languages in Sesotho sa Leboa, or Northern Sotho, results in borrowing, which ultimately leads to an increase in the vocabulary of Sesotho sa Leboa. The languages influencing developments in the vocabulary of Sesotho sa Leboa include, *inter alia*, English, Afrikaans, Xitsonga, Tshivenda and the Nguni languages. This article aims to give a critical analysis of the advantages and disadvantages of the adoption of foreign lexical items in the lemmatization of vocabulary in Sesotho sa Leboa dictionaries. Most puristic inclined academics and intellectuals are reluctant to adopt loan words as a means of developing the Sesotho sa Leboa vocabulary. When confronted with borrowing, the purists usually prefer coinage, using indigenous lexical items to name foreign concepts. This is disadvantageous to the development of the vocabulary of the language because (1) there is no increase in the number of the lexical items in the language since only the meanings of the foreign lexical item are added to existing indigenous lexical items, and (2) in most cases, previously adopted loan words are mistaken for indigenous lexical items and given preference to newly acquired lexical items which have direct and accurate bearing on the meaning of the newly discovered or designed concepts or objects.

Keywords: BORROWING, LOAN WORDS, LEXICAL ITEMS, FOREIGN ACQUISITION, FOREIGN WORDS, COINAGE, COINED LEXICAL ITEMS, PURISM, STANDARD LANGUAGE, STANDARDIZATION, LEMMATIZATION, DEVELOPED LANGUAGES, CORPORA

Opsomming: Ontlening en leenwoorde: Die lemmatisering van nuutverworwe leksikale items in Sesotho sa Leboa. Die invloed van vreemde tale op Sesotho sa Leboa, of Noord-Sotho, het ontlening tot gevolg, wat uiteindelik lei tot 'n toename in die woordeskat van Sesotho sa Leboa. Die tale wat ontwikkelinge in die woordeskat van Sesotho sa Leboa beïnvloed, sluit onder andere Engels, Afrikaans, Xitsonga, Tsivenda en die Ngunitale in. Die doel van hierdie artikel is om 'n kritiese ontleding van die voordele en nadele van die oorneem van vreemde leksikale items by die lemmatisering van die woordeskat in woordeboeke van Sesotho sa Leboa te gee. Die meeste puristies gesinde akademici en intellektuele is onwillig om leenwoorde oor te neem as 'n manier om die woordeskat van Sesotho sa Leboa te ontwikkel. Wanneer hulle met ontlening gekonfronteer word, verkies puriste gewoonlik nuutskeppinge deur gebruikmaking van

* An earlier version of this article was presented as a paper at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the Woordesboek van die Afrikaanse Taal, Stellenbosch, South Africa, 1–3 July 2008.

inheemse leksikale items om vreemde konsepte te benoem. Dit is nadelig vir die ontwikkeling van die woordeskat van die taal omdat (1) daar geen toename in die aantal leksikale items is nie aangesien slegs die betekenis van die vreemde leksikale items tot die bestaande inheemse leksikale items toegevoeg word, en (2) in die meeste gevalle voorheen oorgeneemde leenwoorde aangesien word vir inheemse leksikale items en daaraan voorkeur gegee word bo nuutverworwe leksikale items wat 'n direkte en presiese verband met die betekenis van die nuutontdekte of -ontwerpte konsepte of objekte het.

Slutelwoorde: ONTLENING, LEENWOORDE, LEKSIKALE ITEMS, VREEMDE AANWINSTE, VREEMDE WOORDE, NUUTSKEPPING, NUUTGESKEPTE LEKSIKALE ITEMS, PURISME, STANDAARDTAAL, STANDAARDISERING, LEMMATISERING, ONTWIKKELDE TALE, KORPUSSE

1. Introduction

When compared to most of the highly developed European languages like English and some of the highly advanced African languages such as Kiswahili, Sesotho sa Leboa is still a quite underdeveloped language as far as its lexicon is concerned. Almost all these advanced languages use loan words to develop their vocabulary. English, for instance, absorbs large numbers of foreign words, and this is one of the reasons why the corpora for many indigenous African languages will never be able to compete with the English corpus because its growth rate is always higher and faster than that of most of the African languages. This is due to its rate of acquiring new terminology through borrowing or foreign acquisition. English, for instance, is spoken in many countries on all five continents and is, therefore, able to acquire vocabulary from many languages worldwide. This is coupled with the readiness and the ability of the language to acquire new terminology through borrowing, instead of following the puristic policies.

2. What is borrowing?

Foreign acquisition through borrowing is the adoption of foreign lexical items in a language to refer to new concepts which originated in foreign cultures or foreign languages. Mojela (1991: 12) says:

When people of varied cultures come into contact, they have many things to share and these result in the process of foreign acquisition and an extensive increase in vocabulary. An increase in vocabulary is at the same time accompanied by an increase in meaning.

Gumperz (1968: 223) also alleges the following with regard to borrowing:

Whenever two or more speech communities maintain a prolonged contact within a broad field of communication, there are cross-currents of diffusion.

Borrowing of lexical items from foreign languages occurs directly or indirectly. Mojela (1991: 13) describes direct borrowing in Sesotho sa Leboa as follows:

Words are borrowed from foreign languages and are incorporated into the linguistic system of Northern Sotho.

Direct borrowing results in the formation of loan words, which are subjected to the phonological system of the borrowing language. Higa (1980: 285) emphasizes this fact as follows:

When words are borrowed by one language from another, they are subjected to the phonological rules of the language that borrows.

Indirect borrowing usually occurs through coinage where only the meaning, or the sense of the foreign term is taken, and not the word itself. According to Mojela (1991: 19):

Indirect borrowing is found where a foreign or a new concept is taken over, and not the word itself. In most cases such a concept is associated with an indigenous word. This type of borrowing usually leads to the formation of polysemous words. The word which previously referred to a known concept in Northern Sotho has its meaning extended to refer, also, to the new concept which has closer affinity with the known concept.

These coined words only came into existence after the Sesotho sa Leboa speakers were introduced to these new concepts.

The following are some examples of loan lexical items adopted through direct borrowing, and their coined counterparts, the last two examples being cases of indirect borrowing where existing lexical items have been given extended meanings:

Loan word: *malekere* 'sweets' (from Afrikaans *lekkers*)

Coined word: *dimonamonane* (formed by the reduplication of *mona* 'suck' and the diminutive suffix (*a*)*ne* to express intensity)

Loan word: *radio* 'radio'

Coined word: *seyalemoya* (meaning literally 'that which goes with the wind')

Loan word: *mpete* 'bed'

Coined word: *bolao* 'place to sleep' > 'bed'

Loan word: *tšhinkamo* 'chewing gum'

Coined word: *motu* 'resin' > 'chewing gum'

Foreign acquisition through coinage is usually manifested through the semantic extensions of related indigenous lexical items to include the meanings of the foreign lexical items. This process is sometimes called neologism, which the *Longman Modern English Dictionary* (Watson 1976) refers to as 'the use of words or phrases for a new meaning'.

The main objective of this article is to analyze critically the role played by borrowing in the development of the Sesotho sa Leboa lexicon. The research also analyzes the role played by purism as an impediment to the development of vocabulary in Sesotho sa Leboa.

3. The role played by borrowing in Sesotho sa Leboa

Borrowing plays a major role in the development of languages like Sesotho sa Leboa, especially as far as the following is concerned:

- The development of vocabulary.
- Keeping the language up to date with the latest linguistic developments in all the different spheres of life such as, for instance, the scientific and technological terminologies.
- Increase in the volume of the vocabulary through the conversion of foreign lexical items into loan words in Sesotho sa Leboa.
- In a multilingual society, like South Africa, borrowing helps to bridge the lexical and morphological gaps existing between the various African languages. The meanings of the loan lexical items in Sesotho sa Leboa, for instance, will without difficulties be understood by the speakers of other languages. The loan lexical items in many African languages, which are derived from English, will have more or less the same semantic and morphological features, as shown in the following examples:
 - morphology: *mofolotši* (in Sesotho sa Leboa), and *mūfūloji* (in Ciluba)
 - battery: *peteri* (in Sesotho sa Leboa), and *bhatiri* (in chiShona)

4. Defining a loan word

Mojela (1991: 14) describes a loan word as follows:

A word which has been taken up in the linguistic system of the borrowing language in such a way that it has become part and parcel of the borrowing language.

The loan words in Sesotho sa Leboa conform to its linguistic system. This includes all the phonological, morphological and lexical adaptations of the loan words. For example, the nominal loan words often conform to the nominal class system of Sesotho sa Leboa, e.g.:

peila > *dipeila* 'axe' > 'axes' (from Afrikaans *byl*)
galase > *digalase* 'glass' > 'glasses'
khomphutha > *dikhomphutha* 'computer' > 'computers'

However, class prefixes have to be provided in many cases to accommodate nominal loan words, e.g.:

lebotlelo (class 5) 'bottle'

The Sesotho sa Leboa verbal derivational system applies automatically to the newly acquired loan verbs, i.e.:

bereka > *babereki* 'work' > 'workers'
sebereki > *dibereki* 'hard worker' > 'hard workers'
iperekela 'work for oneself'
mmereko 'a work' or 'a job'
berekela 'work for someone'
berekiša 'cause to work'

5. Borrowing and loan words prior to standardization

Borrowing and the development of loan lexical items prior to the standardization of Sesotho sa Leboa proceeded spontaneously and uninterrupted by common barriers such as linguistic rules, purism and the pressures exerted by the language control and monitoring bodies. Many loan words in most indigenous African languages were formed during the period prior to the commencement of the standardization. The foreign lexical items, derived from languages such as, among others, Afrikaans, Dutch, German, English and the Nguni languages, were used spontaneously by the Sesotho sa Leboa community. There was no controlling or discouraging the use of these foreign lexical items because they helped the indigenous communities to name the new concepts which originated from these foreign cultures. These new lexical items were adopted automatically into the linguistic and phonological system of the indigenous languages. When the standardization of the African languages commenced, the majority of the existing loan lexical items were already in circulation. For instance, at the beginning of the standardization period of Sesotho sa Leboa, most of these previously adopted lexical items were regarded by the majority of the people, including the purists, as indigenous lexical items, since the words were by then incorporated within the linguistic system of Sesotho sa Leboa. Lack of linguistic barriers and puristic approaches during the period prior to the standardization led to an increase in the rate of foreign acquisition, which again resulted in a faster growth of the volume of the Sesotho sa Leboa vocabulary. The following are examples of lexical items acquired during this period, i.e. the period prior to the advent of puristic attitudes and protective linguistic rules in Sesotho sa Leboa:

tšhelete 'money' (Afrikaans *geld*)
manala 'nails' (Afrikaans *naels*)
pere 'horse' (Afrikaans *perd*)

lepolanka 'plank' (Afrikaans *plank*)
lebili 'wheel'
katse 'cat' (Afrikaans *kat*)
lejakane 'Christian' (Afrikaans *diaken* 'deacon')
sekhambeleni 'sweetened water' (Ndebele *skhambeleni* 'what have we travelled for (if not for this)?')
lefamolele or *lefamolebe* 'person who went away for a long time' (Afrikaans *vanmelewe* 'long ago')
puku 'book'
sefepi 'whip' (Afrikaans *sweep*)
tšhofa 'push' (English *shove*)

Most language users (and even purists) did not identify the majority of these lexical items as loan words. They just regarded them as forming part and parcel of the indigenous vocabulary of Sesotho sa Leboa. As a result, many loan words originated during this period, with no one questioning issues such as spelling, consistency and violation of rules.

6.1 Defining purism

Lexical purism can be defined as a policy of excluding foreign lexical elements from the language so that it can remain 'pure'. The *Longman Modern English Dictionary* (Watson 1967) describes purism as 'a strict emphasis on purity, especially in language'. This dictionary further defines a purist as:

a person who places great emphasis or overemphasis on linguistic purity — freedom from foreign words or bastard form. A person who will admit no departures from some chosen method, technique or ideal of perfection.

The following are some of the attitudes and opinions of purists with regard to indigenous African languages:

- They disapprove of foreign influence in the language. The foreign elements are regarded as distorting and contaminating the language, instead of developing and empowering it.
- They prefer the language to develop from within and not from without. Preference is given to indirect borrowing as a positive form of foreign acquisition, in the belief that the indigenous languages have sufficient terminology to serve as variants or synonyms for the new concepts which originate within foreign cultures and languages.
- They regard the indigenous languages as self-sufficient, able to survive without foreign influence.

6.2 The significance of purism in a language

The direct result of purism in Sesotho sa Leboa was the narrow and one-sided standardization which led to the neglecting and, thereby, stigmatizing of most of its indigenous dialects. The majority of the scholars and intellectuals who dominated the membership of the Language Boards came from the few dialects spoken in the areas of Gasekhukhune, Maleuskop and the areas around Polokwane, Mokopane and Waterberg. These areas were the first to have missionary activities, mostly German missionaries, who converted Sesotho sa Leboa (especially Sepedi) into written form. Those who served on the Language Boards standardized Sesotho sa Leboa according to their own dialects and neglected most of the dialects in the Lowveld regions, the Northern and the North Western parts of the former Transvaal. These neglected dialects include those like Sepulana, Selobedu, Setlokwa, Sehananwa and Seroka, i.e. those spoken by the majority of the people who were still poorly educated, owing to a lack of missionary activities in these regions. These scholars and intellectuals play a protective role to ensure that the 'inferior' dialects will not have any influence on the standard language, and that standard Sesotho sa Leboa will remain 'pure'. The same methods which were used to exclude the 'stigmatized' dialects from the standard language are now used to discourage the use of the loan lexical items and to exclude most of these lexical items from standardization, on the grounds that there are suitable indigenous Sesotho sa Leboa lexical items to substitute them.

As a result, the influence of purism in a language, especially an underdeveloped language like Sesotho sa Leboa, usually leads to the following:

- Very slow growth of the vocabulary of the language: Since most of the new inventions originate in (and evolve from) the highly developed and industrialized countries of the world, there will always be a need to adopt vocabulary from these foreign languages. Unfortunately, however, purism will always thwart this process, retarding the progress of the language.
- Foreign acquisition through coinage: Linguistic coinage, which is the only way to foreign acquisition, obviously leads to widespread ambiguity in the language. Most coined lexical items in Sesotho sa Leboa need to be supplemented by a qualificative to disambiguate them, e.g.:

Loan word: *theléfomo* 'telephone' (direct borrowing)
Coined word: *mogala* 'telephone' (indirect borrowing)

The word *theléfomo* refers only to a telephone, while *mogala* refers to (a) a string or rope used to tame heifers, (b) a rope, and (c) a telephone. It is therefore ambiguous. The sentence, *o swere mogala* 'he/she is holding a telephone/rope/string', will consequently not always mean *o swere theléfomo* 'he/she is holding a telephone'.

- As a result, the language will not have enough loan lexical items to name the newly acquired concepts which originate from foreign languages.

7. Conclusion

With regard to the development of the vocabulary of Sesotho sa Leboa, the most important recommendations are the following:

- While supporting the use of coinage or indirect borrowing as a means of foreign acquisition, the use of direct borrowing, which leads to the adoption of loan words or transliterated lexical items, is also an important and accurate method of creating and developing new lexical items in Sesotho sa Leboa.
- The loan words can be used as synonyms to the coined indigenous lexical items.
- Ill-considered purism in Sesotho sa Leboa, which is mostly applied by the indigenous scholars and intellectuals, should be discouraged, because it is disadvantageous to Sesotho sa Leboa, leading to lexical impoverishment.

In contrast to the coined lexical items which are mostly ambiguous, loan words have single meanings, and as such, it will always be easy to lemmatize unambiguous loan lexical items rather than struggling with polysemous and ambiguous coined lexical items.

References

- Gumperz, J.J.** 1968. The Speech Community. Sills, D.L. (Ed.). 1968. *International Encyclopedia of the Social Sciences. Volume 9*: 381-386. New York: The Macmillan Company/The Free Press.
- Higa, M.** 1980. Sociolinguistic Aspects of Word-borrowing. Mackey, W.F. and J. Ornstein (Eds.). 1980. *Sociolinguistics: Studies in Language Contact*: 275-292. The Hague: Mouton.
- Mojela, M.V.** 1991. *Semantic Changes Accompanying Loan Words in the Northern Sotho Lexicon*. Unpublished M.A. Thesis. Pretoria: Vista University.
- Watson, O. (Ed.)**. 1976. *Longman Modern English Dictionary*. Harlow: Longman.

The Effects of Shona Language Change on Monolingual Lexicography: The Need for a Revised Alphabet

William Zivenge, *Department of African Languages and Culture, Midlands State University, Gweru, Zimbabwe (williamzivenge@yahoo.com)*,

Gift Mheta, *Department of Linguistics, University of the Western Cape, Bellville, South Africa (giftmheta@yahoo.com)*

and

Maxwell Kadenge, *Department of Linguistics, University of the Witwatersrand, Johannesburg, South Africa (maxwell.kadenge@wits.ac.za)*

Abstract: In this article, the phenomenon of Shona language change, its effects on lexicography and the need for a revised alphabet are discussed. Because of the defective Shona alphabet, lexicographers at the African Languages Research Institute (ALRI) encountered problems in handling some words that were potential headwords in dictionaries they were compiling under the ALLEX project. The current Shona alphabet is unable to realize quite a number of sounds and morphemes in lexical items in everyday use by the native Shona speakers, because they are alien loanwords. The article was prompted by the challenges encountered during the compilation of *Duramazwi reMimhanzi* (Shona Musical Terms Dictionary). It shows how language change accounts for the problem of headword selection and how modifying the current alphabet can enhance monolingual Shona lexicographical work vis-à-vis the development of the Shona language. It therefore stresses the need for a revised alphabet so as to solve orthographical problems during dictionary compilation.

Keywords: LANGUAGE CHANGE, ALLEX, LEXICON, SEGMENT, SUPRASEGMENT, ALPHABET, LEXICOGRAPHY, ASSIMILATION, BORROWING, ADOPTION, ARTICULATION, SOUNDS

Pfupiso Yechinyorwa: Matambudziko Anosanganwa Nawo Mukugadzira Maduramazwi eMutauro Mumwe chete neKuda Kwekushanduka Kwemutauro weChiShona: Panofanirwa Kuva neManyorerwo eChiShona Matsva.

Muchinyorwa chino vanyori vanoongorora kushanduka kuri kuita mutauro weChiShona neku-famba kuri kuita nguva nematambudziko anosanganikwa nawo nekuda kweshanduko iyi zviku-rusei mukugadzira maduramazwi ari mururimi rweChiShona chete. Vanyori vanoonesawo pfungwa yekuti panodiwa manyorerwo matsva. Nyanzvi dzinogadzira maduramazwi paAfrican Languages Research Institute (ALRI) dzinosangana namadambudziko esarudzo yamazwi padzi-ninge dzichisarudza mazwi adzinenge dzichida kuisa mumaduramazwi adzinogadzira pachi-

rongwa chokugadzira maduramazwi chinonzi ALLEX. Dambudziko resarudzo yemazwi rinowanikwa nekuda kwechimiro chebumbiro remanyorerwo eChiShona risingatenderi mamwe mavara ari mune imwe mitauro asiri muChiShona. Bumbiro iri parizvino haritenderi mamwe mavara, mibatanzidzwa yamavara namamwe mazwi anowanikwa mumutauro wemazuva ano waVaShona anobva mune mimwe mitauro. Chinyorwa chino chakatokonywa namadambudziko akasanganiwa nawo nevanyori vebepa rino pavakange vachigadzira *Duramazwi reMimhanzi*. Chinyorwa chino chakanangana nokuburitsa kushanduka kuri kuita ChiShona, uye matambudziko anosanganwa nawo pakusarudza mazwi anopinda muduramazwi uye kuonesa zvakare kuti kugadzirisa bumbiro ramanyorerwo kunogona kurerutsa basa ravagadziri vamaduramazwi. Izvi zvinoita zvakare kuti mutauro weChiShona uvandutswe. Pamusoro pezvose izvi, chinyorwa chino chinoda kutaridza kuti sei zvakakosha kuita chimbichimbi kugadzirisa chimiro chebumbiro remanyorerwo pakuzama kurerusa matambudziko anosanganwa nawo nevagadziri vamaduramazwi eChiShona.

Mazwi Akakosha: KUSANDUKA KWEMUTAURO, ALLEX, MAZWI, SEGIMENDI, SUPURASEGIMENDI, ARIFABETI, ZVAMADURAMAZWI, ADHAPUTESHENI, KUKWERETA, ADHOPUSHENI, ATIKURESHENI, MITINHIRO

Introduction

Linguists, especially Ferdinand de Saussure (Atchison 2001: 18), have shown that change is inevitable in any language. Languages change through various ways, the commonest being adopting and assimilating segmental and supra-segmental features from languages with which they are in contact. Through this contact, languages constantly interact as they communicate objects within their linguistic environments. Speakers of such languages mingle and may exchange linguistic items, which ultimately interferes with their mother tongues, thereby altering them. They may part with some of their linguistic aspects and acquire new ones, adding them to their inventories. Some languages signify objects that are peculiar to themselves but, because of constant interaction with other languages, natural transfer occurs as speakers mingle, resulting in what Chimhundu (2002) refers to as adoption. This means that change is, among other reasons, the result of borrowing linguistic features from one language into another to fill in communication gaps in the receiving language. Usually languages borrow segments from foreign languages with which they are in contact and/or from other indigenous languages or dialects. Contact with these other languages necessitates cross-linguistic influence as a result of political, cultural, social and economic developments in the world. Bynon (1977) describes language contact as the existence of more than one language in a particular locality. Shona is directly in contact with English as a foreign language, which came into Zimbabwe owing to colonization, but Shona also mingles with distant foreign languages via other indigenous languages. Shona, comprising Karanga, Manyika, Zezuru, Korekore and Ndau dialects, is one of the two widely spoken languages in Zimbabwe. It is spoken by people who

mostly live in the eastern half of Zimbabwe, adjacent to Mozambique. It is in contact with English, Ndebele and a wide range of community languages.

Apart from English, Shona is also in contact with fourteen other indigenous languages that have segments different from those of Shona. Some of the popular ones are Shangani, Ndebele, Nambya, Tsonga/Tonga and Venda. It is from some of these languages that Shona obtains some of its lexical items to add to its lexical inventory. In other words, it receives additional linguistic features from outside and from within its traditional domain. In this process of borrowing, from outside its main domain, there is a high chance that it will incorporate alien segments and suprasegments into its inventory. Cultural intimacy and dialect borrowing for the past decades have resulted in new segments and suprasegments in Shona. Some of the dialects of Shona such as Ndau, Manyika and Korekore are in border areas. These border areas are points of ethnolinguistic enclave. Speakers of these dialects are living in contact with speakers of other languages found in Mozambique. Because of this the Manyika and Ndau people become bilingual, speaking their neighbours' languages. Such ethnic interaction results in the exchange of linguistic features, which results in Shona receiving other sounds from outside its domain.

Invasion and the subsequent colonization by the British in the 1890s were followed by periods of linguistic contact between English and Shona. This was conducive to the diffusion of lexical items between the two languages to such an extent that most Shona speakers have become bilingual. The fact that English was then elevated to an official language, used in educational and formal sectors, gave it a prestigious status. This became a conditional motivating factor to acquire English as a way of asserting affiliation with a prestigious language. This linguistic legacy, even into the present day, coerces Shona speakers to be associated with English, which entails speaking it or borrowing from it. Education also as a social factor has actually prepared Shona speakers to acquire English for it has been made the medium of instruction in schools, colleges and universities. This means that all those who have passed through formal education to any level, have acquired English to some degree. This means that three-quarters of Shona-speaking youth have gone through formal tutelage where English has been used.

Migration to and from Zimbabwe by the Shona and speakers of other languages for economic, political and social reasons, which increased after independence in 1980, has created complex linguistic contact between Shona and other languages, causing an extensive transfer of linguistic elements. The spread of international languages has also been greatly made possible and influenced by the electronic media, in particular computer, video and audio technologies. In this situation linguistic segments are being transferred from one language to the other, Shona being no exception.

In this process languages gain or lose particular linguistic and extralinguistic attributes and so does Shona. In situations where a language is in constant change, especially where new sounds, morphemes, lexical items and

suprasegmental features are adopted from other languages which do not share similar features, lexicographers face serious problems in handling headwords with such characteristics. As established in this article, a wide range of changes occur in Shona because of its contact with English and other African languages, which ultimately results in Shona adopting some characteristics of these languages and then adapting them to or assimilating them in Shona.

Patterns of change

Shona is developing through the borrowing of lexical items to fill in gaps realized in communication. This ultimately results in sound and phonological changes. Other changes are realized at the suprasegmental level. At the segmental level, change is mediated by phonetic processes so as to ease articulation. A language borrows because of a lack or need realized by gaps in communication, which are referred to as lexical gaps. Shona disseminates (sounds and morphemes) to and borrows (sounds and lexical items) from other languages, thereby losing some of its properties and gaining others, which Chimhundu (2002) refers to as transfer of elements.

These new sounds that do not correspond with the current orthography then pose problems to lexicographical work, particularly treating them as entries in dictionaries. It is within the framework of this article to look at Shona language change at the level of sound and above. In most cases when words are borrowed from another language which do not share similar sounds with the receiving one, assimilation takes place to accommodate those words with alien sounds in the new phonological environment. In most of the sounds complete assimilation is achieved, especially where there is direct correspondence between sounds of the loaner language and Shona. Usually the phonological conventions available in Shona will accommodate them as shown, for instance, by the following correspondences:

| English | Shona | Word in Shona with English gloss | |
|---------|-------|----------------------------------|-------|
| m | m | <i>mita</i> (meter) | m > m |
| t | t | <i>mita</i> (meter) | t > t |
| s | s | <i>sofa</i> (sofa) | s > s |
| c | k | <i>kapu</i> (cup) | k > k |
| p | p | <i>kapu</i> (cup) | p > p |

Such circumstances do not effectuate any graphological changes, the concept (signified) is adopted but the sounds are nativised or lexicalized through phonological assimilation. The ultimate result of this rephonologization process is what is termed complete assimilation. Loanwords such as *mita* (meter), *sofa* (sofa) and *kapu* (cup) can be accommodated by the Shona orthography and they do not register any graphological changes.

There are situations where there are no direct equivalents or where there are no sounds near to the Shona ones. In complex situations where there are no direct correspondences, alien sounds from the loaner language are adopted as they are into the Shona language as shown below.

| English | Shona |
|----------------|--------------------------------------|
| theory | <i>thiyori</i> |
| thermometer | <i>themomita</i> or <i>themometa</i> |
| thousand | <i>thausendi</i> |

The postulated changes in these loanwords from English are that the voiceless inter-dental fricative [θ] has been incorporated into Shona in the borrowing process. This is what may be referred to as partial assimilation. What this implies is that there is a blend of sounds found in the loaner language and receiving language, for instance, in the word *thiyori* (theory), the word-initial [θ] is English whilst *yori* is Shona. Other examples of English consonant sounds that have been assimilated in Shona but are not represented in Shona orthography are the English lateral approximant [l] and the breathy voiced alveolar trill [ɾ] as shown in the words below.

| English | Shona |
|----------------|-------------------|
| lotion | <i>losheni</i> |
| loaf | <i>lofu</i> |
| lecture | <i>lekicha</i> |
| library | <i>laibhurari</i> |
| lorry | <i>rhor</i> |
| ruler | <i>rhula</i> |

The problem that compilers of dictionaries face is whether to include such words in the Shona dictionary or not. What has been explained so far is consonant assimilation in the process of nativisation or what Chimhundu (2002) refers to as adaptation. This shows that consonant assimilation may be partial, resulting in the incorporation of new sounds and a segmental change.

The above discussion has therefore brought the realization that Shona has new sounds for which there is no provision in Shona orthography, for Shona has only the following sounds, from which various acceptable combinations are derived:

<a, b, bh, c, ch, d, dh, e, f, g, h, i, j, k, m, mh, n, nh, ng, ny, n', o, p, r, s, sh, sv, t, u, v, vh, w, y, z, zh, zv>

The monolingual lexicographer's postulated challenges

A monolingual Shona lexicographer is envisaged to face a major challenge of headword selection in so far as orthographic plausibility is concerned. The se-

lection of some headwords having new sounds adopted from other languages was problematic in *Duramazwi Guru ReChiShona* (DGC) and especially in *Duramazwi ReMimhanzi* (SMTD). These are words like *thiyori* (theory), *thiyeta* (theatre), *loni* (lawn), *rhori* (lorry), *themometa* (thermometer), *rheza* (razor) and *lita* (litre).

All these words which exist in the Shona corpus, are widely used by Shona speakers. As they were borrowed from English, their spellings are not recognized in Shona. The challenge is whether to include or exclude them from the dictionary. Excluding them would imply that they are not commonly used, which is untrue as evidenced by their existence in the corpus. Including them bring with it the problem of presentation, that is, whether to represent their spoken form with the letters and letter combinations permissible in the Shona alphabet or to follow the orthography but misrepresent the spoken form. The dilemma of the lexicographer is whether or not to treat them as headwords in monolingual Shona lexicography, with the purist users of both DGC and SMTD strongly feeling that their language can be corrupted by entering English-spelt words not fully realized through Shona orthography.

There are also other words from different dialects of Shona that present a similar problem to monolingual Shona lexicographers. Magwa (2002) says:

From 1967 onwards, speakers of different dialects were experiencing certain difficulties arising from the defective alphabet and the spelling and word division system. The current orthography is linguistically constricting in a number of ways. For example, the standard alphabet does not have symbols representing the sounds <l> and <x>, which are found in ChiKaranga, ChiNdau and ChiKorekore dialects.

What can be deduced from this discussion is that some words borrowed from English, Nguni and Xhosa cannot be realized by the Shona. Examples of such words are:

| Word | Dialect in which it is used | English gloss | Problem letters/letter combinations |
|-----------------|-----------------------------|-----------------|-------------------------------------|
| <i>muhlobo</i> | Karanga/Ndau | way of doing | -hl-, -b- |
| <i>pxere</i> | Karanga | kids | -x-, -px- |
| <i>xumbudzi</i> | Karanga/Ndau | lean goat | -hx- |
| <i>muhlaba</i> | Ndau | jocular/naughty | -l-, -hl- |
| <i>maxeu</i> | Karanga | maheu | -x- |
| <i>xarani</i> | Karanga | thread | x- |
| <i>muxoro</i> | Karanga | wage/salary | -x- |
| <i>kudla</i> | Karanga | eat | -dl- |
| <i>ntunzwi</i> | Ndau | fly | nt- |
| <i>khamba</i> | Ndau | leopard | kh- |
| <i>nquzu</i> | Ndau | type of dance | -q-, nq- |

These words exist in the Shona corpus, some of them appearing quite frequently, but were not treated as headwords in DGC, owing to the fact that the current orthography does not recognize them. However, some of these words, both from English and other African languages, were treated as headwords but have changed spellings as shown below:

pxere > *pwere* (kids)
uxwa > *uswa* (thatching grass)
muhlobo > *mutovo* (way of doing)
muxoro > *muhoru* (salary)
lita > **lita* (litre drink)
themometa > **themometa*

The asterisk shows that there is a violation of the orthography, whilst all the other examples without an asterisk are a misrepresentation of the spoken forms. Language change has brought with it challenges to monolingual lexicographers at the African Languages Research Institute (ALRI).

The challenges that lexicographers are facing need an urgent revision of the Shona alphabet so that it becomes flexible to accommodate all patterns of language change. Accommodating change is feasible only when the alphabet allows a diversity of sounds and sound combinations. No language can develop if change is resisted or remains unrecognized.

Conclusion

The discussion has undertaken to show how Shona is developing by adopting lexical items from other languages with which it is in contact. This results in the incorporation of some sounds alien to Shona. These are basically those sounds that cannot be realized by the current Shona alphabet. The article has indicated the problems of headword selection arising as a result of the inadequacies of the current orthography. The article concentrates only on problems emanating from English loanwords and those borrowed from other African languages. The emphasis of this article has been to analyze how headword selection has become problematic by different word categories as a result of Shona language change. This is an important aspect to consider for monolingual lexicography because technical evolution of the twenty-first century, with the advent of globalization, is causing both linguistic and orthographic evolution, a shift that is evidenced by the inadequacies of the Shona orthography. What is important, however, is that monolingual lexicographers, together with both speakers and planners of the language, should constantly revise and broaden the alphabet and orthography of their language, to cater for language development. This will help to overcome the problem of the inadequacies of the orthography.

Bibliography

- Aitchison, J.** 2001. *Language Change: Progress or Decay?* Third Edition. Cambridge: Cambridge University Press.
- ALLEX.** 1999. *Advanced Shona Dictionary Style Manual*. Unpublished Document. Harare: University of Zimbabwe.
- Benson, T.G.** 1964. A Century of Bantu Lexicography. Guthrie M. (Ed.). 1964. *African Language Studies V. Collected Papers in Oriental and African Studies*: 65-91. London: University of London.
- Bynon, T.** 1977. *Historical Linguistics*. Cambridge: Cambridge University Press.
- Chabata, E.** 1998. Using the Predictability Criterion for Selecting Extended Verbs for Shona Dictionaries. *Lexikos* 8: 140-153.
- Chimhundu, H.** 1994. Zimbabwe Language Situation. Asher, R.E. (Ed.). 1994. *The Encyclopedia of Language and Linguistics. Volume 9*: 5084-5085. Oxford/New York: Pergamon Press.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Chimhundu, H.** 2002. *Adoption and Adaptation in Shona*. Oslo: Unipub Skriftserier.
- Chitauro-Mawema, M.B.** 2002. Challenges Encountered in the Compilation of an Advanced Shona Dictionary. *Lexikos* 10: 209-224.
- Gumperz, J.J.** 1970. Sociolinguistics and Communication in Small Groups. Pride, J.B. and J. Holmes (Eds.). 1972. *Sociolinguistics: Selected Readings*: 203-224. Harmondsworth: Penguin.
- Magwa, W.** 2002. The Shona Writing System: An Analysis of its Problems and Possible Solutions. *Zambezia* 29(1): 1-11.
- Mheta, G. (Ed.)**. 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Tulloch, S.** 1994. *Reader's Digest: Complete Word Finder*. Oxford: Reader's Digest.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. Janua Linguarum. Series Maior 39. The Hague: Mouton.

More than Twenty Years of Svensén's *Handbook*

Loránd-Levente Pálfi, *Centre for Lexicography, Aarhus School of Business,
University of Aarhus, Denmark (llp@asb.dk)*

Abstract: The number of real handbooks dealing with lexicography and metalexicography is not, and has never been, impressive. When one looks at comprehensive handbooks and in addition those brought up to date, the number becomes even smaller. Because of this, but also because it has existed for more than 20 years (i.e. in different editions/versions) and because it is — especially in Scandinavia — a quite well-known work, Svensén's handbook is an important publication. In this review article, I will consider what the critics have written over the years (i.e. on the different editions/versions of Svensén's handbook), and I will attempt to determine whether the book is about to become a classic.

Keywords: LEXICOGRAPHY, METALEXICOGRAPHY, HANDBOOKS, SCHOLARLY REVIEWING, CITATIONS, *A HANDBOOK OF LEXICOGRAPHY*, BO SVENSÉN, *THE ART AND CRAFT OF LEXICOGRAPHY*, SIDNEY I. LANDAU, *MANUAL OF LEXICOGRAPHY*, LADISLAV ZGUSTA

Opsomming: Meer as twintig jaar van Svensén se handboek. Die aantal werklike handboeke wat die leksikografie en die metaleksikografie behandel, is nie en was nog nooit indrukwekkend nie. Wanneer 'n mens na omvattende handboeke kyk en daarby dié wat bygewerk is, word die aantal selfs kleiner. Om dié rede, maar ook omdat dit vir meer as 20 jaar bestaan (d.w.s. in verskillende uitgawes/bewerkings) en omdat dit — veral in Skandinawië — 'n redelik welbekende werk is, is Svensén se handboek 'n belangrike publikasie. In hierdie resensieartikel sal ek oorweeg wat die kritici oor die jare geskryf het (d.w.s. oor die verskillende uitgawes/bewerkings van Svensén se handboek), en sal ek probeer bepaal of die boek besig is om 'n klassieke werk te word.

Sleutelwoorde: LEKSIKOGRAFIE, METALEKSIKOGRAFIE, HANDBOEKE, VAKKUNDIGE RESENSERING, AANHALINGS, *A HANDBOOK OF LEXICOGRAPHY*, BO SVENSÉN, *THE ART AND CRAFT OF LEXICOGRAPHY*, SIDNEY I. LANDAU, *MANUAL OF LEXICOGRAPHY*, LADISLAV ZGUSTA

The book discussed in this article is Bo Svensén's *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making* (2009). A previous English version was published under the title *Practical Lexicography. Principles and Methods of Dictionary-Making* (i.e. Svensén 1993). Originally the book appeared in Swedish in 1987. A revised and extended edition of the Swedish version was published under the title *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik* (i.e. Svensén 2004).

Whereas, on the one hand, the difference between *Practical Lexicography* [...] (1993) and *Handbok i lexikografi* [...] (2004), as Varantola (2006: 107) also partly argues, and, on the other hand, the difference between *Practical Lexicography* [...] (1993) and *A Handbook of Lexicography* [...] (2009), as Trap-Jensen (2010: 223) argues, is big enough for the latter two, i.e. the titles from 2004 and 2009, to be regarded as new versions rather than new editions compared with the titles from 1993 and 1987, the difference between *Handbok i lexikografi* [...] (2004) and *A Handbook of Lexicography* [...] (2009) is quite small — apart from the fact that the one is written in Swedish while the other is a somewhat revised edition written in English; cf. however Trap-Jensen (2010: 226):

The English edition [i.e. Svensén 2009] is much more than a translation of the Swedish original [i.e. Svensén 2004]. Many content revisions have been made to cover the developments that have taken place in the five-year period between the two editions. The most obvious difference, however, is the large number of examples that have been replaced to cater for the change in readership from Scandinavian users to the English-speaking international world. The Swedish examples used for monolingual dictionaries have largely been replaced by English ones, and bilingual examples are usually between English as L1 and French or German as L2. Similarly, the bibliographical references have been adapted to an international audience, dominated by English texts with a supplement of German and French texts and only a few selected texts in a Scandinavian language.

Whether new versions or new editions, Svensén's handbook has been on the market for more than 20 years. And all the editions/versions have been produced by Svensén himself (except the version from 1993, which is a translation by two professional translators). This is a remarkable achievement by one man in the modern history of metalexigraphy. It seems to be time to make a status report on the life of this handbook. In the following, I will take a look at what previous reviewers have commented on the work (the versions of 1993 and 2004), and I will try to establish whether or not the critique has resulted in changes in the English edition from 2009. I will also look at critique of the 2009 edition, adding my own comments. There is only one review of Svensén's handbook that is not treated with here: A very short review written by Ladislav Zgusta and published in the DSN journal *Dictionaries* 14, dealing with the first Swedish version from 1987, and, because it forms part of a discussion of several other books, containing little communication on the book.

McCreary (1997) is the first reviewer to be dealt with here. He reviews Svensén (1993) and thereby raises the following "minor irritations" (McCreary 1997: 346): In his opinion the book (a) has a Eurocentric perspective, (b) is written in a dry prose, and (c) has an inadequate index. Besides these problems, McCreary (1997: 347) is quite positive on the book.

Regarding point (a), he says:

As a lexicographer with experience in Japanese–English dictionaries, I was disappointed with the book's exclusively Eurocentric perspective. However, the

chapters above on the construction of definitions and the choice of bilingual equivalents could be helpful to the lexicographer in Asia (or anywhere for that matter) who wants to examine the various methods of definition construction and the various orderings of senses for polysemous words.

I do not agree with McCreary on this issue. One can take a look on this matter from an opposite angle, as Varantola (2006: 108) does:

A clear advantage of the Handbok is that Svensén is extremely well-read in multilingual lexicographical literature and is thus able to describe lexicographical practices in different countries. In addition to English, he also draws examples from a number of other languages, such as Swedish, German and French. In this way he is able to highlight problems that do not necessarily come up in English dictionaries.

Regarding McCreary's (1997: 347-348) point (b), the following two quotations can serve as clarification:

In a number of chapters, the reader has a feeling that the English has been closely and accurately translated, but is lacking in idiomatic natural English, which unfortunately lends a dry tone to the work. Some suitably colorful English idioms along with a livelier writing style would have been appreciated. [...] Considering that it is an exhaustive treatment of all of the elements in dictionaries and systematically presents all of their functions for both native and non-native users, the text is very readable, although not as readable as Landau's (1984) work. [...] Svensén's book is similar to Zgusta's (1971) text. Similarly, this quality of systematic thoroughness can make it a suitable text for a graduate course in lexicography. However, those who choose it for a graduate course might want to supplement it with another more general book with a more readable style, such as Landau's (1984) book.

In this regard, I must agree with McCreary — and I can add that this has not changed for the better in the 2009 edition. On the contrary. The 1993 version was translated by two professional translators; the 2009 edition has been translated by Svensén himself. This is my opinion; others may disagree, as for instance Trap-Jensen (2010: 226):

Combined with the fact that the English translation from Swedish is, to my judgment, accurate and idiomatic, conscientiously rendering the author's precise and sober-minded style of writing, there is no reason, neither with respect to focus nor to language, why the book should not find the large and responsive audience it deserves.

Regarding McCreary's point (c), the following quotation can serve as clarification: "The text concludes with a useful but occasionally cursory subject index (pp. 281-285), which should have been supplemented by an author index" (McCreary 1997: 347). In this case, there have been some improvements: The index

in the 2009 edition is much more adequate than the one in the 1993 version, but an author index is still lacking. And even if the index in the 2009 edition is better than that found in many other books, it still has some shortcomings; cf. for instance the entry "L2→L1 dictionaries":

L2→L1 dictionaries (*see also* bilingual dictionaries)
14-18, 37, 61-2, 66, 68, 70, 73, 88, 115, 125, 126,
128, 129, 131, 144, 155-7, 175-7, 181, 193, 198,
202-3, 260, 263, 267, 269-71, 275-6, 278-9, 293,
299, 313, 324, 349, 353-7, 376, 393, 395-6, 446,
464-5, 466, 467-9

In the field of LIS (i.e. library and information science), this is called a burial — an LSP word denoting that so many references are given that the information, which one wishes to make accessible, in fact literally becomes buried, and therefore inaccessible. However, in many other connections Svensén is obviously aware of the concept of burials, since he often makes the necessary specifications by using sub-entries (which, by the way, is the common solution for this problem); cf. for instance the entry "lexicography":

lexicography
bilingual 46, 55-7, 61-2
bottom-up 449-50
computerization of 401, 402, 410, 414, 437
corpus 45-58, 330, 404, 415-20
corpus-based *see* corpus lexicography
'corpus-driven' 58
definition of 2-3
documentary 1, 65
ethical aspects of 9, 427, 431-2, 486
general-language 3, 220, 227
...

The index certainly deserves to be praised. At the same time, there are still quite a number of burials that should have been avoided. And it is regrettable that the reader still cannot look up names.

The next two international critics in line are Varantola (2006), who reviews Svensén (2004), and Trap-Jensen (2010), who reviews Svensén (2009). The former (Varantola 2006: 107) is of the opinion, that Svensén should have given much more space to electronic lexicography:

Only in one chapter is the electronic medium in all its dimensions treated separately, namely in the section dealing with electronic dictionaries. [...] the electronic medium has affected so many lexicographical axioms and principles of dictionary compilation including entry presentations that they alone would merit a new standard handbook in the field.

This has not changed in the 2009 edition. Cf. Trap-Jensen (2010: 226), who obviously agrees with Varantola (2006) and repeats the critique in somewhat other words:

Even though the author has no doubt invested considerable effort to incorporate many new aspects pertaining to e-dictionaries throughout the book, the book cannot avoid a bias towards the printed medium. The next generation of lexicographic textbooks will, at least in this reviewer's opinion, have to reflect the two-sided nature of modern dictionary-making: on the one hand, it must deal with the compilation, organization and storage of dictionary data in a database independently of the channel of publication and, on the other hand, it must discuss how those data are presented to the user, whether in a printed book or on a screen. That would lead to the inclusion of chapters on database architecture, data markup and schemas and it would lead to other chapters dealing with topics such as web design and user customization for e-dictionaries, and space-saving strategies such as the use of abbreviations and symbols for print dictionaries. And in return, chapters on micro-, macro- and cross-reference structure are more or less irrelevant for e-dictionaries and would have to be rewritten to be compatible with the dual model of data organization and data presentation.

One should keep in mind, however, that the world is not spinning around the web. The number of printed media is bigger today than it has ever been before. Billions of documents, actually, are being produced in the world each day (cf. Pálfi 2010b). The web is not included in this number. Surely a new handbook (or several handbooks) with the main focus on electronic lexicography and the digital world must — and will probably soon — appear. Svensén's handbook with its main focus on the printed media will still be necessary.

The next critical point from Varantola (2006: 107) regards the targeted readership and the overall style (macro- and megastructure) of the book:

The scope of the manual is broad and therefore problematic, because it is clear that these target groups [i.e. lexicographers, language students and teachers, language planners, and translators] have different needs and interests. It is difficult to imagine that the book could as such be used as a textbook for language students as it is too detailed in its lexicographical classifications. On the other hand, the book could be gainfully used as a reference work and additional reading material for lexicographical projects in language studies.

Trap-Jensen (2010: 224) has a somewhat opposite view on this topic (it should be noted that Trap-Jensen's review regards the 2009 edition, whereas Varantola's review regards the 2004 version):

A connected question is whether the book is primarily an introduction to lexicography as an academic discipline or a reference book for practitioners in the field. The book itself does not provide the answer. [...] The structure of the book is such that each chapter addresses one component of the dictionary entry and analyses it in relation to the various types of dictionaries. This is the result of a choice made by the author and it has the advantage that all the components can

be described exhaustively in their own chapters and by the use of comparative analysis. In my judgment, this structure is more expedient in a textbook for metalexigraphers and students of lexicography but less attractive when used in a handbook for dictionary practitioners. The latter are not likely to be interested in comparative analyses of how the same dictionary element is treated in different dictionaries. More likely, they want to learn what kind of data they should include in their dictionary and how to present it in the concrete project at hand. Imagine a lexicographer writing a bilingual L2–L1 dictionary: she would have to read through the whole chapter on the relevant element, carefully picking out those passages that are relevant for her L2–L1 dictionary. This makes it more difficult for her to use the book as a handbook for quick reference in everyday situations. The impatient reader who wants a quick overview of a specific problem will have to try his luck via the initial table of contents or the 15 pages subject index at the end of the book.

I do not agree with either of the two. Svensén's book is a typical, classical handbook in the European tradition which goes back for centuries. This kind of book provides in-depth information — the incomparable Hausmann et al. (1989–1991) also is such a kind of handbook — and is not meant to be used for quick reference like a dictionary. If in-depth information on the core aspects of lexicography cannot be obtained from a handbook like this, where then should one look? Should one read several monographs (instead of reading chapters comprising about 10–30 pages in Svensén's handbook) every time one wants to gain insight into a specific aspect of lexicography or metalexigraphy, i.e. an insight much more comprehensive than that gained by a regular dictionary look-up (for instance in a dictionary like Martínez de Sousa 1995, Burkhanov 1998 or Hartmann and James 2006)? The targeted readership of handbooks of this type is broad, so this is characteristic of this kind of publication. I find it problematic to present — as an example of a potential user — a superficial, impatient dictionary-compiler (cf. the quote from Trap-Jensen above). Compiling a dictionary is not a task with which one is occupied for a week or two. One must do thorough and serious research, which includes not only reading Svensén's handbook or consulting some chapters in it, but also reading many other sources.

Varantola (2006: 108) is complimentary on the bibliography: "The book [...] has a very extensive bibliography, a treasure for anyone embarking on a study of any aspect of lexicography, as well as a detailed index." Trap-Jensen (2010: 225) seems to agree, as he mentions "the voluminous 32 pages bibliography at the end of the book". Svensén's bibliography is not thin, but it does not impress either if compared with bibliographies in other publications. Cf. for example Considine (2008), which is a monograph consisting of 322 pages (bibliography and index excluded) and containing a bibliography comprising 52 pages. (See also Pálfi 2009b: 511.) Considine's bibliography is indeed impressive considering that it is a monograph, and not a handbook. Another example: The bibliography in Pálfi (2010a) comprises 172 pages (the whole book consisting of 432 pages). It is a dictionary of dictionaries similar to Kabdebo and

Armstrong (1997), the difference being that the dictionary of Kabdebo and Armstrong concerns English dictionaries (i.e. dictionaries from the UK, the USA, Australia, Ireland etc., with English as either language of communication or as one of the languages treated within the dictionaries in question), while Pálfi (2010a) concerns Danish dictionaries. A praiseworthy bibliographic feature in Svensén's handbook (the 2004 version and the 2009 edition), however, is, according to Trap-Jensen (2010: 225-226), the following:

As a new feature, the book has at the end of each chapter a section with suggested further reading, not just in the form of a list of titles and authors but an annotated list with specifications on the subject, on model examples or on theoretical discussions of a particular aspect. [...] The value of this feature can hardly be exaggerated and will undoubtedly save a lot of time and effort for lexicographers and students on the lookout for relevant literature.

Another praiseworthy feature of the book, on which also Varantola (2006: 108) comments regarding the 2004 version, and which has been left unchanged in the 2009 edition, is the following:

Svensén is a very non-committal writer. He does not normally advocate one approach or solution over another but describes different solutions and principles impartially. Only towards the end of the book, where he tackles ethical and legal issues, such as plagiarism, copyright, and marketing ethics, does he take firmer standpoints. In other words, Svensén has obviously wanted to write a manual that provides multi-faceted information about the various dimensions of dictionaries and dictionary-planning for the benefit of lexicographers and editors who need the information for their own decisions.

The reverse side of this, however, is the dry prose/tone, which is probably also connected with the non-committal writing style (however, the Swedish version of 2004 does not suffer from this dry prose/tone). All in all, both Varantola (2006) and Trap-Jensen (2010) are very positive about Svensén's handbook. At the same time, the latter raises a criticism, with which I agree:

A single point of criticism stands out, however: the book is bound to be more or less out of date from the very day of its appearance. This is of course often the case with publications in a rapidly changing field and cannot be helped. Unfortunately, the situation is in this case aggravated by the time spent on translation and the adaptations needed for international publication: Svensén began working on the Swedish edition some years before it appeared in 2004 and it took another 5 years before the current English edition came out.

Admittedly this is a choice not easily made: Either one can share the work with one or several other scholars (thereby speeding up the time of production of the given book) and possibly also safeguard oneself against errors of detail, that otherwise might pass unnoticed (cf. Pálfi 2009c: 518), or one can choose to be the sole author. Also co-authorship has its negative consequences: the risk of lack of terminological consistency, of cohesion between the individual chap-

ters, of common basic assumptions and starting-points, of common theoretical groundwork etc. (cf. Pálfi 2009c: 518-519).

The longest and most thorough review of Svensén's handbook comes from Vikør (2005), who deals with Svensén (2004). Vikør's review which comprises 16 pages is written in Norwegian. Being initially and conclusively very positive, Vikør partly raises a number of errors of detail in his critique, and partly he is trying to summarize the contents of the book. Only some of the most significant errors of detail observed by Vikør shall be mentioned in the following. For instance, Vikør (2005: 338) points out that Svensén (2004: 28) erroneously regards synonym dictionaries as onomasiological dictionaries, which, Vikør argues, is a misunderstanding (although one which not only Svensén is guilty of, but also numerous other scholars in the field of lexicography and metalexicography), since synonym dictionaries have an alphabetical macrostructure — opposed to thesauruses, that have an onomasiological macrostructure, i.e. according to meaning. This has been corrected in the 2009 edition, cf. for instance Svensén (2009: 30-31). Vikør (2005: 336-337) furthermore considers Svensén's (2004: 29-30) treatment of the dictionary and the norm (i.e. normative vs. descriptive dictionaries etc.) inadequate. This also seems to have been changed in the 2009 edition (cf. for instance Svensén 2009: 24). Vikør (2005: 334-335) finds it somewhat problematic that the Swedish version of 2004 focuses too strongly on Swedish lexicography: The book would have been more Nordic in essence and concept had it also contained more examples from Norwegian and Danish dictionaries, i.e. more references to Norwegian and Danish lexicography (not to speak of Finnish, Icelandic, Faroese, and Greenlandic lexicography). However, this detail does not relate to the English edition of 2009, which is solely conceived for an international audience, and in which most of the examples have been replaced by mostly English material.

The only entirely negative reviewer in the existence of Svensén's handbook is Tarp (2005). His review is written in Danish. Although Tarp also makes some positive remarks, he is initially, medially and conclusively negative. For instance, Tarp (2005: 6) highlights some introductory passages from Svensén (2004), which in content have not changed in the 2009 edition: "There are those who argue that lexicography is a kind of applied linguistics, while others (among them the author of this book) regard it as in independent discipline" (Svensén 2009: 3). The problem here, Tarp argues, is that this formal position of Svensén does not affect the rest of the book. Tarp subscribes to this position that lexicography/metalexicography is a separate scholarly discipline and not a part of linguistics (cf. for instance Tarp 2008 *et passim*). Svensén obviously agrees, but, according to Tarp, does not take the consequences of this position further than page 3 (in the Swedish version of 2004 page 4). Another example highlighted by Tarp is the following (Svensén 2009: 2):

Lexicography is an activity which consists in observing, collecting, selecting, analyzing and describing, in a dictionary, a number of lexical items (words, word elements and word combinations) belonging to one or more languages.

In the Swedish version of 2004, this paragraph appears on page 3. Tarp (2005: 6) argues: That which makes lexicography a separate scholarly discipline in relation to linguistics is the fact that it has dictionaries and not language as its theoretical focus, and that its practical focus is the making of dictionaries for a given user group in accordance with optimizing the genuine function of the dictionary as a tool for information retrieval. Hence it is not "lexical items" that is the pivotal point, but those specific problems that specific user-groups may encounter in specific user situations. In addition, Tarp (2005: 7) argues that Svensén's handbook cannot and should not be used to learn something about lexicographic theory. The dictionary typology proposed in Svensén (2004) Tarp also finds quite problematic, since it relies more on traditional linguistic scholarship than on modern metalexicography. However, the pages on dictionary typology have been rewritten in the English edition of 2009. It therefore seems Svensén has accepted some of the criticism raised by Tarp. Another problem of concern, Tarp argues, is the chapter 'Ordboksprojekt' (in the English edition of 2009 it is called 'Dictionary projects'). Here, among others, Svensén distinguishes between "project plan" and "dictionary plan" and gives the reader the impression that the former is mainly conceived by administrators, marketers, economists and legal consultants, while the latter is conceived by the lexicographer(s) only (this goes for the Swedish version of 2004 as well as the English edition of 2009; for the latter, see for instance pages 399-402 and 405-406). Tarp finds this more than problematic, since the "project plan" in Svensén's chapter includes such very central issues as dictionary functions: It is mainly the lexicographer, Tarp argues, who should occupy him-/herself with such issues as dictionary functions, user groups and user needs, not administrators, marketers, economists or legal consultants. I agree with this specific aspect. As becomes clear from above, these paragraphs do not seem to have been changed or corrected in the 2009 edition. All in all, Tarp argues that the book should be used by specialists, i.e. lexicographers and metalexicographers only, because these will be likely to see the gaps and inadequacies. Students and semi-experts (among these specialists from neighbouring disciplines) should not use the book, i.e. in the view of Tarp. To accommodate Tarp's critique completely would have meant rewriting extensive parts of the book — which Svensén has not done with the 2009 edition. Even though, for a greater part, I agree with Tarp, I find the issue not as problematic as he does. For some reason or another, Svensén still seems to have managed to write a truly metalexicographic book opposed to, for instance, Considine (2008), which is a book from and for the field of LIS (i.e. library and information science) rather than from and for the field of lexicography and metalexicography. What I concluded about Considine (2008) (cf. Pálfi 2009b: 514), certainly does not apply to Svensén's handbook:

The most significant and recurring deficiency of the book is the general lack of metalexicographic reflection. [...] *Dictionaries in Early Modern Europe* is evidently written by a philologist with a profound interest in historical bibliography for other philologists with the same interest. [...] The book will certainly be of great

value and use primarily for philologists, librarians, book historians, historical bibliographers, and secondarily for historians and cultural historians. Dictionary historians, in so far as these regard themselves metalexigraphers, and metalexigraphers in general, will find the book lacking in metalexigraphic reflection. [Pálfi 2009b: 512] All in all, *Dictionaries in Early Modern Europe* is an exceptionally erudite, thorough and trustworthy book [...], however, the book is of little value for metalexigraphers.

Svensén's handbook is about to become a classic — in the sense of 'a publication that is being re-edited and reprinted again and again, and is known and used around the world'. Or is it? In the following, I will look at some citation figures. The figures for IJL were harvested with the help of the search engine on the IJL website (however, one has to be aware of the fact that the total number given for a search does not equal the right information: one has to check every search result for every publication found by the search engine to exclude reviews of the given books, self-citations, double showings, etc.). The figures for *LexicoNordica* have been harvested manually, since *LexicoNordica*, although it has a homepage and an online as well as a printed version, does not have a search engine like IJL.

One can only wonder why Svensén's handbook (1987/1993/2004) is cited surprisingly little in comparison with Zgusta's *Manual of Lexicography* (1971), which, one would think, is rather outdated. Thus Svensén's handbook (1987/1993/2004) has until now (i.e. June 2010) only been cited 12 times in *International Journal of Lexicography*, the first time being in 1998. The three citations by the reviewers McCreary (1997), Varantola (2006), and Trap-Jensen (2010) are not included in this figure. From these 12 scholars¹ only one is from the USA, while the 11 others are from Europe (8), China (1), Japan (1), and Australia (1). From these 12 citations, 10 refer to *Practical Lexicography* [...] (1993), and two to *Handbok i lexikografi* (2004). Zgusta's *Manual of Lexicography* (1971) is cited 29 times (in this figure the two occurring self-citations from Zgusta himself are not included). The citations are by scholars from Europe (13), the USA (7), Japan (3), Russia (1), China (1) and Australia (1), while three citations are of mixed origin.² If one looks at Sidney Landau's *Dictionaries: The Art and Craft of Lexicography* (1984/1989/2001), the following figures appear: Europe (16),³ the USA (5), Japan (3),⁴ Canada (2),⁵ China (2), New Zealand (2), Kuwait (1), Jordan (1),⁶ Australia (1),⁷ while two citations are of mixed origin.⁸ The figures for Landau are 35 in total. The two reviews of Landau's book in IJL and the two occurring self-citations by Landau himself are not included in these figures. That IJL was launched in 1988 compensates for the fact that Landau's book first appeared in 1984, whereas the first English version of Svensén's book was published in 1993. The figures for the three are as follows (citations are included from any kind of contribution, i.e. articles, review articles and reviews of other books than Zgusta 1971, Landau 1984/1989/2001, and Svensén 1987/1993/2004 investigated here):

| Landau, Zgusta, and Svensén | Citations in IJL 1988–June 2010 |
|--|---------------------------------|
| <i>Dictionaries</i> [...] (1984/1989/2001) | 35 |
| <i>Manual of Lexicography</i> (1971) | 29 |
| Svensén's handbook (1987/1993/2004) | 12 |

If one considers the citation figures for Zgusta (1971), Landau (1984/1989/2001), and Svensén (1987/1993/2004) in the journal *LexicoNordica* (which is a scholarly, peer-reviewed journal of about 250-350 pages, appearing annually in November since 1994 as a publication of the Nordisk Forening for Leksikografi, cf. Pálfi 2009a), the following results can be presented (as above, with the same exceptions, citations being included from any kind of contribution):

| Svensén, Landau, and Zgusta | Citations in <i>LexicoNordica</i> 1994–2009 |
|--|---|
| Svensén's handbook (1987/1993/2004) | 43 |
| <i>Dictionaries</i> [...] (1984/1989/2001) | 5 |
| <i>Manual of Lexicography</i> (1971) | 5 |

Svensén (1987/1993/2004) is cited 43 times in total: Swedish (24),⁹ Danish (10),¹⁰ and Norwegian (9).¹¹ In these figures Vikør (2005), which is the only review of Svensén's handbook in *LexicoNordica*, is not included, nor are six self-citations by Svensén. For Landau, the specifications are as follows: Swedish (4),¹² and Danish (1). The only review of Landau (2001) in *LexicoNordica* is not included in these figures. For Zgusta, the specifications are as follows: Swedish (3),¹³ and Danish (2).

The information on the nationality of the different contributors to IJL is taken from the contributions themselves (i.e. the information on the country of residence), as these are always given (in reviews at the end, and in articles at the beginning). This is not the case with *LexicoNordica*. Only in recent years has it become common practice to give information about the contributors at the end of each contribution, therefore languages and not countries have been chosen above for contributors to *LexicoNordica*. However, the languages used in *LexicoNordica* do not equal the nationality of the contributors: For instance, Finnish scholars always publish in Swedish (mostly in Finno-Swedish, i.e. Swedish as spoken and written in Finland), even though it is allowed to publish in Finnish, whereas Icelandic, Faroese and Greenlandic scholars usually publish in Danish (some Icelandic scholars also publish in Norwegian), because most scholars in Scandinavia understand Swedish, Norwegian Bokmål, and Danish, but only a few understand Finnish, Icelandic, Faroese and Greenlandic (i.e. Inuit). In addition: All Finno-Swedish contributions are regarded as Swedish contributions in the above discussion.

Even though it is debatable what a classic is, the citation figures shown above combined with the fact that Svensén's handbook seems to be a steady seller, indeed points in the direction that Svensén's handbook is about to become a classic; even if not in the world generally, certainly in Scandinavia specifically.

Endnotes

1. The number of scholars is actually 14, because two of the relevant contributions are written by two scholars each; however, the number of citations is 12 equalling the 12 contributions.
2. Mixed origin: One of the IJL contributions is written by three scholars (the one from the USA, the other from Spain, and the third from The Netherlands); two of the IJL contributions are written by one scholar working in Europe (Belgium) as well as in Africa (South Africa); the Australian contribution is written by two scholars. Two of the contributions actually appeared in the EURALEX Newsletter included in IJL, and not in IJL itself: One of these contributors is from the USA, the other is from Europe.
3. Four of the European contributions citing Landau's book are written by two different European scholars each (i.e. eight scholars in total).
4. One of the three Japanese contributions citing Landau's book is written by five Japanese scholars.
5. One of the two Canadian contributions citing Landau's book is written by two Canadian scholars.
6. The Jordanian contribution citing Landau's book is written by two Jordanian scholars.
7. The Australian contribution citing Landau's book is written by two Australian scholars.
8. These two contributions citing Landau's book are written by the same scholar working in Europe (Belgium) as well as in Africa (South Africa).
9. One of the Swedish contributions citing Svensén's book is written by two Swedish scholars.
10. Two of the Danish contributions citing Svensén's book are written by two Danish scholars each (i.e. four scholars in total); and one of these contributions cites two different versions (i.e. Svensén 1993 and 2004).
11. Three of the Norwegian contributions citing Svensén's book are written by two Norwegian scholars each (i.e. six scholars in total).
12. One of the Swedish contributions citing Landau's book is a translation (by someone else than the author) of a review article originally written in German, for the purpose of being published in *LexicoNordica*; whether the German version has been published as well, is not clear from the information given in a note at the end of the contribution; information is only given on the fact that it has been translated from a German original.
13. One of the Swedish contributions citing Zgusta's book is a translation of a review article originally written in German (see note 12 above).

References

- Burkhanov, I.** 1998. *Lexicography: A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Considine, J.** 2008. *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. Cambridge: Cambridge University Press.
- Hartmann, R.R.K. and G. James.** 2006. *Dictionary of Lexicography*. London/New York: Routledge. [First Edition 1998].
- Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.).** 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexico-*

- graphy/Dictionnaires. Encyclopédie internationale de lexicographie. Volume 1–3.* Berlin/New York: Walter de Gruyter.
- Kabdebo, T. and N. Armstrong.** 1997. *Dictionary of Dictionaries and Eminent Encyclopedias: Comprising Dictionaries, Encyclopedias, and other Selected Wordbooks in English.* Assisted by Etain Ó Síocháin. Second Edition. London: Bowker-Saur. [First Edition 1992].
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography.* First Edition. New York: Charles Scribner's Sons.
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography.* Reprint of the First Edition. Cambridge/New York: Cambridge University Press.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography.* Second Edition. Cambridge/New York: Cambridge University Press.
- Martínez de Sousa, J.** 1995. *Diccionario de lexicografía práctica.* Barcelona: Bibliograf.
- McCreary, D.R.** 1997. [Review of] Bo Svensén: *Practical Lexicography. Principles and Methods of Dictionary-Making.* Translated from the Swedish by John Sykes and Kerstin Schofield. Oxford/New York: Oxford University Press 1993. *International Journal of Lexicography* 10(4): 346-348.
- Pálfi, L.-L.** 2009a. LexicoNordica. EURALEX Newsletter, Summer 2009. *International Journal of Lexicography* 22(2): 234-235.
- Pálfi, L.-L.** 2009b. [Review of] John Considine: *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage.* Cambridge: Cambridge University Press 2008. *Lexikos* 19: 509-515.
- Pálfi, L.-L.** 2009c. [Review of] A.P. Cowie (Ed.): *The Oxford History of English Lexicography. Volume I: General-purpose Dictionaries. Volume II: Specialized Dictionaries.* Oxford: Oxford University Press 2009. *Lexikos* 19: 516-523.
- Pálfi, L.-L.** 2010a. *Leksikon over ordbøger og leksika.* Assisted by J.F. Lindholm. Copenhagen: Frydenlund.
- Pálfi, L.-L.** 2010b. Det digitale hysteri — eller kunsten at synge i kor. *BogMarkedet* 9: 25.
- Svensén, B.** 1987. *Handbok i lexikografi. Principer och metoder i ordboksarbetet.* Stockholm: Esselte Studium & Tekniska Nomenklaturcentralen.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making.* Translated from the Swedish by John Sykes and Kerstin Schofield. Oxford/New York: Oxford University Press.
- Svensén, B.** 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik.* Second, Revised and Extended Edition. Stockholm: Norstedts Akademiska Förlag.
- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.
- Tarp, S.** 2005. Anmeldelse af Bo Svensén: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik.* Andra, omarbetade och utökade upplagan. Stockholm: Norstedts Akademiska Förlag 2004. *LEDA-Nyt* 39: 5-8.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Lexicographica. Series Maior 134. Tübingen: Max Niemeyer Verlag.
- Trap-Jensen, L.** 2010. [Review of] Bo Svensén: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press 2009. *International Journal of Lexicography* 23(2): 223-227.

- Varantola, K.** 2006. [Review of] Bo Svensén: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Second, Revised and Extended Edition. Stockholm: Norstedts Akademiska Förlag 2004. *International Journal of Lexicography* 19(1): 107-110.
- Vikør, L.S.** 2005. Ei metaleksikografisk perle. Bo Svensén: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Andra, omarbetade och utökade upplagan. Stockholm: Norstedts Akademiska Förlag 2004. *LexicoNordica* 12: 331-346.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton.
- Zgusta, L.** 1992/93. [Review of] Bo Svensén: *Handbok i lexikografi. Principer och metoder i ordboksarbetet*. Stockholm: Esselte Studium & Tekniska Nomenklaturcentralen 1987. *Dictionaries* 14: 176-178.

H.A. Welker and Pedagogical Lexicography

Sven Tarp, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, South Africa, and Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (st@asb.dk)*

Abstract: The lexicographer Herbert Andreas Welker's book of more than 500 pages about pedagogical lexicography provides a panoramic overview of the various types of pedagogical dictionaries as well as the corresponding theoretical literature. Welker's method is descriptive, with himself mainly an observer of lexicographical practice and a "collector" of — frequently opposed — opinions and ideas expressed by a large number of scholars. This method allows the reader to become acquainted with an important part of the most relevant literature on pedagogical lexicography which is presented in a systematic and condensed form. Written in Portuguese, the book might perhaps not be so easily accessible to many readers. This review article provides a guided tour through the main contents of this highly recommendable book and discusses some of the most important ideas reproduced in it.

Keywords: PEDAGOGICAL LEXICOGRAPHY, PEDAGOGICAL DICTIONARIES, LEARNERS' LEXICOGRAPHY, LEARNERS' DICTIONARIES, SCHOOL DICTIONARIES, CHILDREN'S DICTIONARIES, DESK DICTIONARIES, COLLEGE DICTIONARIES, DICTIONARIES FOR FOREIGN-LANGUAGE LEARNERS, DICTIONARIES FOR MOTHER-TONGUE LEARNERS, DICTIONARIES FOR LEARNERS OF SCIENTIFIC DISCIPLINES

Opsomming: **H.A. Welker en die pedagogiese leksikografie.** Die leksikograaf Herbert Andreas Welker se boek van meer as 500 bladsye oor die pedagogiese leksikografie verskaf 'n panoramiese oorsig oor die verskillende soorte pedagogiese woordeboeke sowel as die ooreenstemmende teoretiese literatuur. Welker se metode is beskrywend, met homself hoofsaaklik 'n waarnemer van die leksikografiese praktyk en 'n "versamelaar" van — dikwels teenstellende — menings en opvattinge uitgespreek deur 'n groot aantal vakkundiges. Hierdie metode laat die leser toe om kennis te maak met 'n belangrike deel van die tersaaklike literatuur oor die pedagogiese leksikografie wat in 'n sistematiese en saamgevatte vorm aangebied word. Omdat dit in Portugees geskryf is, kan die boek miskien nie so maklik toeganklik wees vir baie lesers nie. Hierdie resensie-artikel verskaf 'n begeleide reis deur die hoofinhoud van hierdie hoogs aanbevelenswaardige boek en bespreek 'n aantal van die belangrikste opvattinge wat daarin weergegee word.

Sleutelwoorde: PEDAGOGIESE LEKSIKOGRAFIE, PEDAGOGIESE WOORDEBOEKE, AANLEERDERSLEKSIKOGRAFIE, AANLEERDERSWOORDEBOEKE, SKOOLWOORDEBOEKE, KINDERWOORDEBOEKE, HANDWOORDEBOEKE, WOORDEBOEKE VIR VREEMDETAAL-AANLEERDERS, WOORDEBOEKE VIR MOEDERTAALLEERDERS, WOORDEBOEKE VIR LEERDERS VAN WETENSKAPLIKE DISCIPLINES

1. Introduction

Herbert Andreas Welker is a German lexicographer with residence in Brasília, Brazil. He has produced three books and several articles with reflections on various aspects of lexicography. An important part of this production is written in Portuguese, a fact that reflects his interest in contributing to the theoretical and practical development of lexicography in Brazil where, on the one hand, still relatively few researchers master English or German at an academic level and where, on the other hand, the last decade has seen an increase in the production of dictionaries, especially for school purposes, owing to government policy and priorities. Unfortunately, this also impedes most lexicographers outside the Portuguese-speaking and -reading community from becoming acquainted with his work, especially the three important books written during the last few years: Welker (2004) which is a short introduction to lexicography; Welker (2006) which is the first and most ambitious panoramic work on lexicographical user research ever published (see reviews by Lew 2007 and Tarp 2008a), and Welker (2008) which, like its predecessor, provides a comprehensive, and so far unprecedented, view of an important and relevant part of lexicography, i.e. pedagogical lexicography.

Like the previous works, this last book — the *Panorama geral da lexicografia pedagógica* (General Survey of Pedagogical Lexicography) — is mainly descriptive. Welker provides an extensive panorama of current lexicographical practice as well as of theoretical reflections by a large number of scholars, with himself in the role of an observer and a "collector" without contributing many new ideas. This modest approach is justified when it comes to informing the Brazilian lexicographical community about the international state-of-the-art of lexicography. However, for an "outsider" — who is not part of the main target group of readers — it would also have been useful to receive more theoretical input from the author himself. This, however, cannot really detract from a meritorious work which in all respects is worth reading.

The book consists of seven main chapters with the following titles:

- (1) What is pedagogical lexicography?
- (2) A short history of pedagogical lexicography
- (3) On language learners' use of dictionaries
- (4) Dictionaries for learners of foreign languages
- (5) Native-language pedagogical dictionaries
- (6) Special pedagogical dictionaries
- (7) Electronic dictionaries

In the following sections, the contents of these chapters will be discussed consecutively. All quotations are English translations of the original Portuguese text.

2. What is pedagogical lexicography?

This chapter, probably the most interesting from a typological point of view, is also the chapter where the author's personal opinion is expressed more distinctly. Welker starts with a brief overview of the different positions regarding the academic status of lexicography (see Welker 2004: 11 and Tarp 2010), making it clear that he himself, contrary to Wiegand (1984), uses the terms *theoretical lexicography* (or *metalexigraphy*) and *practical lexicography* to denote the theoretical and practical aspects of *lexicography*, a term which he accepts as a hyperonym of the two other terms. In this respect Welker reveals a point of view which may not be surprising for those who are well-acquainted with lexicographical literature, but which is nevertheless averse to a discipline trying to gain academic status, i.e. that these terms are used by the various scholars with a lot of different meanings and frequently without providing any definition.

According to Welker, the term *pedagogical lexicography* is seldom used; for instance, it is not included in any of the titles in the *International Encyclopedia of Lexicography* (Hausmann et al. 1989–1991), in the *Diccionario de lexicografía práctica* (Martínez de Sousa 1995) or in various well-known introductions to lexicography, e.g. Landau (2001). Welker is surprised that various books and articles, which do include the terms *pedagogical lexicography* or *pedagogical dictionaries* in their respective titles, only deal with dictionaries for foreign-language learning. This is, for instance, the case with Bogaards (1991), Rundell (1998), Wiegand (1998, 2002), Dziemianko (2006), and Xatara and Humblé (2006). On the other hand, Welker quotes various authors like Hartmann and James (1998), Dolezal and McCreary (1999) and Hernandez (1998) who also include dictionaries for native-language learners within the concept of pedagogical lexicography. He therefore writes (p. 18):

It seems that pedagogical lexicography is more embracing than what many of the contributions quoted above let you think; in fact, it could be said that pedagogical lexicography includes dictionaries conceived for learners of both foreign languages and the mother tongue.

We will later discuss whether this definition is sufficiently broad. Before doing so, however, it is relevant to introduce other elements which Welker uses to encompass the concept of pedagogical lexicography. In this respect, he rejects the idea that teaching dictionary usage is part of pedagogical lexicography (p. 19):

Some people say that teaching the usage of dictionaries is pedagogical lexicography but the latter should not be confused with the pedagogy — or didactics — of dictionary usage.

For Welker (p. 19), a dictionary is a tool produced with a specific purpose in mind:

Sometimes it is claimed that pedagogical dictionaries 'teach' languages, but this is an imprecise statement. Obviously, pedagogical dictionaries do not teach languages ... What they do — or try to do — is to assist the learning of a (foreign or native) language ... As a conclusion, you can say that theoretical pedagogical lexicography (or pedagogical metalexigraphy) studies problems related to pedagogical dictionaries and that practical pedagogical lexicography produces such dictionaries.

In this vein Welker proceeds to define what is meant by the term *pedagogical* in relation to lexicography and dictionaries. Initially he notes that some authors, among them Hernandez (1998), prefer the term *didactic* instead of *pedagogical*. Welker rejects this idea. In his understanding (p. 21), pedagogy is the "theory and science of teaching and education", whereas:

Didactics is a part of pedagogy, but not the same. Didactics is a set of methods and techniques used in order to obtain an efficient teaching.

On this basis, he concludes (p. 21-22):

Pedagogical dictionaries aim at being more didactic. In lexicography, the adjective *pedagogic* refers to a specific *type* of dictionaries (dictionaries conceived to language learners), whereas *didactics* should only be used in order to qualify the *way* in which the information is provided: more or less didactic, more or less clear, more or less adapted to the users' skills. In this respect, even pedagogical dictionaries may vary in their didactic quality.

Welker then continues the discussion on lexicographical terminology. He proceeds from the sphere of theory to that of practice when he discusses the various terms used to denominate pedagogical dictionaries in various languages (Portuguese, Spanish, English, German and French). Without going into detail regarding the various more or less successful translations, there are two fundamental terms used as titles in a large number of pedagogical dictionaries: *school dictionaries* and *learners' dictionaries*. Welker relates these two terms to two different types of pedagogical dictionaries, i.e. those for learners of a native language and those for learners of a foreign language. He writes (p. 22-24):

It seems to me that the term *dicionário escolar* is broadly accepted, both in Portuguese and Spanish. And the same applies to German (*Schulwörterbücher*), French (*dictionnaires scolaires*) and English (*school dictionaries*) ... As to the term *learners' dictionaries*, it is so deep-rooted that everybody knows that it is not used for native-language learners.

Previously, Welker has criticized Hartmann and James (1998: 107) who — in their *Dictionary of Lexicography* — writes that the "distinction usually made between a dictionary for native speakers (school dictionary) and the one for non-native learners (learner's dictionary) is not helpful". Welker (p. 19) disagrees:

On the contrary, I think that this distinction is very useful and necessary.

The problem here is that there seems to be two parallel discussions. The first is whether a distinction should be made between a pedagogical dictionary for native speakers and one for foreign-language learners. It is obvious that the two types of potential users have different types of lexicographical needs. Welker is therefore quite right when he opposes Hartmann and James' statement in this regard. However, another discussion concerns which terms should be used to determine the two types of dictionaries. In this respect a distinction should be made between the precise scientific terms needed in lexicographical theory and the commercial terms used to sell the dictionaries.

The terms *school dictionary* and *learners' dictionary* are not precise terms. The term *school dictionary* implicitly refers to all dictionaries conceived to be used in the school system. This is the way the term is used commercially in a number of countries, e.g. South Africa, where one can buy bilingual school dictionaries, such as the *Tweetalige Skoolwoordeboek Afrikaans–Engels/Engels–Afrikaans*, which are obviously not produced to assist the learning of a mother tongue, and even specialized school dictionaries such as Longman's *Multilingual Science Dictionary for South African Schools* and *Illustrated Dictionary of Natural Sciences and Technology Today* whose purpose is basically to assist the learning of science, not of language. On the back cover of the latter it is explicitly stated:

It enhances conceptual understanding of key concepts and will help learners succeed in Natural Sciences and Technology.

And the same applies to the term *learners' dictionary*, where it is necessary to take into consideration that the word *learner* today is widely used in a much broader way than just to refer to learners of a foreign language (or even of a native language). In fact, an article by Cowie (1996), to which Welker also briefly refers on page 340, has the title "*The 'Dizionario Scolastico': a Learner's Dictionary for Native Speakers*". Likewise, the subtitle of another South African pedagogical dictionary clearly states that it is designed for "*learners, students and trainees in science and technology*" (cf. Hartmann-Petersen et al. 2001), a fact showing that neither the term *pedagogical dictionary* nor the term *learners' dictionary* can be restricted to dictionaries aimed at assisting language learning, whether foreign or native.

The commercial publishing houses cannot be expected to provide scientifically correct titles to all their lexicographical products, especially not when specific terms have already taken root among the public in specific countries. But when it comes to theoretical work, it is necessary to build upon well-defined concepts and use terms with a logical linguistic relation to their content. In this respect, the state-of-the-art of practical lexicography shows that dictionaries are produced to assist not only language learning but also learning of scientific disciplines.

The only logical conclusion is therefore that all these dictionaries — i.e. dictionaries assisting native and foreign language learning as well as scientific

learning — should be considered *pedagogical dictionaries* and that *pedagogical lexicography* should be defined as that part of lexicography dealing with these dictionaries. As the word *learner* is broadly used to denominate not only foreign-language learners, but also mother-tongue learners and learners of scientific disciplines, i.e. the potential users of pedagogical dictionaries as they have just been defined, the terms *learners' lexicography* and *learners' dictionary* should therefore be considered as synonymous with the terms *pedagogical lexicography* and *pedagogical dictionary*.

This conclusion may seem strange after so many years of terminological confusion, but it is, as mentioned, the only logical one. Correspondingly, the term *school dictionary* should be assigned to all types of dictionaries designed for use in the school system to assist in the learning of the mother tongue, a foreign language, or a scientific discipline. In this way, *school dictionary*, like *children's dictionary*, *desk dictionary*, etc., should be used as a subcategory of the category *pedagogical dictionaries* (learners' dictionaries). When this reasoning is followed, the pedagogical dictionary types which Welker calls "school dictionary" and "learners' dictionary" should therefore — at least in the theoretical literature — be renamed and given much more scientifically correct names like *dictionary for mother-tongue learners* and *dictionary for foreign-language learners*, to which should be added *dictionary for learners of scientific disciplines*. Except for the last term, these terms are in fact the ones Welker himself uses in the titles of various chapters in his book. As such, he has taken a big step in the right direction. To this should be added that Welker in a later chapter (p. 113) also recognizes that pedagogical dictionaries may be used for learners of scientific disciplines although he includes this cognitive function in the dictionaries for language learners without considering that it may also give rise to a separate type of pedagogical dictionaries, i.e. dictionaries for learners of scientific disciplines, in cases where it is the only or the main function of these dictionaries.

3. A short history of pedagogical lexicography

This short chapter of only 12 pages consists of two parts, the first dealing with "ancient dictionaries used by language learners" and the second treating the "history of theoretical pedagogical lexicography". In the first part, Welker refers to studies carried out by other scholars and quotes contributions containing the words *learner*, *pupil* and *student*. In this way, an interesting picture is taking shape as it seems that pedagogical dictionaries, especially for language learning, have existed since the very dawn of lexicography. This is, for instance, the case with the several thousand years old Sumerian dictionaries — or "proto-dictionaries" — which were carved in clay and apparently invented as tools to support the training and teaching of future scribes and also with the Sumerian-Assyrian dictionaries which were used to assist Babylonian students who had reception problems when reading Sumerian texts almost 3 000 years ago. In fact, Zöfgen (1994: 289) claims that:

From the beginning, the history of bilingual dictionaries reveals a strong didactic orientation where 'the main motive behind lexicography until 1600 was to help learners of foreign languages'.

Welker then briefly discusses some pedagogical dictionaries produced in Europe in the 16th and 17th centuries. Two of these dictionaries are of special interest. The first one is a bilingual dictionary produced in 1521 to assist Englishmen who wanted to learn French. To those who today oppose the production of monofunctional dictionaries — and even deny their existence — the title of this dictionary tells its own story: *The Introductory to Wryte and to Pronounce French*. The second dictionary worth mentioning is the *Orbis sensualium pictus*, published by Johann Amos Comenius in 1658 and later translated into more than 20 languages. This interesting dictionary does not have access to the relevant data through a central word list, but provides on the left side of the pages a series of pictures with numbers placed in several parts and followed by two text columns where the numbers are repeated: the first column contains Latin words, collocations and even phrases, and the second column their translations into German. The major merit of this 350 years old dictionary was that students could use it to discover hitherto unknown phenomena and assist in the learning of the corresponding vocabulary, in one language or the other. In this way, the *Orbis sensualium pictus* contradicts the frequent linguistic claim that the function of dictionaries is always to define words.

In the second part of the chapter, Welker briefly deals with some of the most important scholars and their main contributions in the field of pedagogical lexicography. Not without reason, he begins in 1899 with the English philologist Henry Sweet who many researchers consider the father of modern pedagogical lexicography. He then continues with the North American linguist and educator Edward L. Thorndike, the British lexicographers, Harold Palmer, Michael West and A.S. Hornby, the Russian Lev Ščerba, the Czech Ladislav Zgusta, the Germans Franz Josef Hausmann and Herbert Ernst Wiegand, the Frenchmen Jean Dubois and Robert Galisson, the Spaniards Humberto Hernandez Hernandez and Maria del Carmen Avila Martin, the Danes Hans-Peder Kromann, Theis Riiber and Poul Rosbach, the Belgians Jean Binon, Thierry Selva and Serge Verlinde, ending with Sven Tarp.

This panoramic overview seems relevant and justified, showing that pedagogical lexicography, in its various expressions, is common to different cultures and epochs. In this respect, Welker could perfectly well have mentioned other scholars from other language communities who in one way or another have contributed to the development of pedagogical lexicography, e.g. the Swede Martin Gellerstam, the architect behind the LEXIN project for immigrant learners of Swedish (cf. Pálfi and Tarp 2009). However, it is interesting that Welker does not mention any Portuguese or Brazilian lexicographers in this chapter, although they are frequently quoted in the remaining part of the book. This fact alone seems to justify the publication of his book in Portuguese.

4. On language learners' use of dictionaries

This chapter is one of the most interesting in the book. The first part reflects the opinions a large number of researchers express about the usage of dictionaries for language learning. Although mainly descriptive, it furnishes a condensed and highly informative panorama of the various opinions and tendencies within pedagogical lexicography since 1955, i.e. during more than half a century. Many present-day lexicographers will be surprised to see that the theoretical and practical problems they are trying to solve today have already been raised and discussed several decades ago. It is impossible to reflect all these ideas in this review article, but at least some of them deserve to be mentioned, for instance Mathews (1955: 187) who appears surprisingly "modern" when he writes:

Dictionaries are tools, and they are much more complicated, and capable of many more uses than students suspect. All of us know students need encouragement and guidance in the use of dictionaries, and perhaps there are few teachers of freshman composition but that devote a part of their program to an effort to help students form the habit of consulting dictionaries.

Other researchers, especially from the 1970s onwards, are much more hesitant when it comes to recommend the use of dictionaries in language learning. Beattie (1973: 163), for instance, writes:

Dictionaries and glossaries ought to be resorted to only *after* 'intelligent guessing' has failed. The reason why dictionaries are harmful is that learners use them *before* attempting an intelligent guess, and consequently never learn how to guess intelligently.

Beattie therefore recommends that learners should not possess dictionaries during the first two or three years, but that they may sometimes consult them in school to become acquainted with their use. Several other authors express similar critical ideas concerning the use of dictionaries in language learning.

On the other hand, Herbst (1985) shows that maybe 30 or 40 percent of all mistakes made by German secondary-school students in English text production could have been avoided if they had consulted bilingual or monolingual dictionaries. He provides three reasons to explain this problem: The students do not consult dictionaries, because (a) they do not know that they can find the needed information in them, (b) they do not know how to interpret the information provided, and (c) they do not realise that they have made a mistake. Herbst therefore recommends the use of dictionaries. This is also the case with Gu (2003) who distances himself from the widespread guessing and mnemonics of the previous decades and seems to come close to the opinion expressed by Mathews almost fifty years earlier:

It is alarming to see how much time and effort we have spent in areas such as contextual guessing or mnemonics and yet how little energy is dedicated to an area such as dictionary strategies that can be just as illuminating.

Welker himself (p. 77-80) also seems to share this position:

There are situations where inference impedes the comprehension — maybe not of the general idea of the text but, for instance, of a whole paragraph — and guessing frequently does not permit an exact understanding of the sentence where the word appears. (...) The conclusion is that it is often not possible to infer the meaning of an unknown word and that it is necessary to consult a dictionary.

As already indicated, these quotations, however interesting, do not do justice to a text which is rich in — frequently opposing — ideas and inspiring for any pedagogical lexicographer interested in improving his/her work, whether theoretical or practical. Other important discussions are also reflected in this chapter through a number of quotations and references, e.g. whether to use monolingual or bilingual dictionaries in foreign-language learning, how to study and assimilate vocabulary and grammar by means of dictionaries, etc. In the rest of the chapter, the observant reader may also find relevant information about lexicographical user research, consultation skills, the teaching of dictionary use, and "the dictionary as didactic material used to acquire non-linguistic knowledge".

5. Dictionaries for learners of foreign languages

This chapter of 178 pages is by far the longest in the book. It discusses already published dictionaries for learners of foreign languages and is divided into two sections, one about monolingual dictionaries and the other about bilingual ones. The *first section*, the longer of the two, is further subdivided into four parts: (a) dictionaries for non-native learners of English, (b) dictionaries for non-native learners of other languages (specifically French, Portuguese, German, Spanish and Italian), (c) characteristics and components of pedagogical dictionaries, and (d) empirical research into the usage of these dictionaries.

In the first part, Welker provides a condensed story of well-known British learners' dictionaries for non-native learners of English, starting with the very first attempts in the early 20th century and ending up with the current Big Five (Oxford, Longman, Collins Cobuild, Cambridge and Macmillan) and the sharp competition between them. In this respect, he also analyzes the three fundamental ideas behind this tradition, i.e. the *controlled vocabulary*, the *verb patterns*, and the *collocations*, as well as the various levels in which the Big Five have been published so far: *advanced*, *intermediate* and *elementary*. Those who have read Cowie (1999) will find few surprises in this chapter, but it is nevertheless informative and recommendable as a short introduction to the subject owing to its condensed, well-written and didactic character.

In the second part, Welker looks at the dictionaries for non-native learners of French, Portuguese, German, Spanish and Italian. Surprisingly, this part is a little shorter than the previous one, a fact which may reflect the number of

contributions published on learners' dictionaries for each of these languages, but which is nevertheless not completely justified, because these language communities, also have a long and interesting history of pedagogical dictionaries for foreign-language learners which, in some respects, provide pedagogical solutions different from those of their English counterparts. Welker's treatment of the dictionaries in these five languages is inspiring, because it provides a wider dimension to an academic field which in many respects is constrained by the not always justified British predominance.

The third part provides a 30 pages long systematic introduction to the most important characteristics and components of monolingual dictionaries for foreign-language learners: *size, outside matter, macrostructure, lemma selection, access, layout, pronunciation, hyphenation, definitions, grammatical information, labels, collocations, examples, idiomatic expressions, illustrations, usage notes*, etc. Although many of these phenomena can be found in other types of dictionaries, there is little doubt that this section constitutes an easy introduction to some of the most important components, data types and characteristics of dictionaries for foreign-language learners. The only problem is that the various phenomena are not presented and discussed in the light of the different lexicographical functions, a common problem in much of the theoretical literature about this type of dictionaries — and other dictionaries as well. This deficiency is probably due to Welker's descriptive approach but it is nevertheless a pity because the author may be one of the few lexicographers who could have combined lexicographical functions and dictionary characteristics and components in a convincing way. Welker himself (p. 26) writes:

It is well-known that publishing houses and authors of dictionaries generally strive to make their products useful in any user situation. However, the ideal would be that every pedagogical dictionary clarifies its main function or functions (reception, production, vocabulary learning) ...

The last part of this section contains references to a number of research projects conducted to analyze the use of dictionaries for foreign-language learners. The 40 pages long text furnishes much interesting material which, regrettably, shares a general problem with the majority of the lexicographical user-research projects carried out so far, i.e. that it does not live up to the standards required by modern sociology and statistics (cf. Tarp 2009). The inevitable result is that the material provided by this sort of research may be interesting and even thought-provoking, but in no way statistically significant, for it cannot be generalized and used for solid theory-building. Of course this cannot be blamed on Welker, but it would have been more convincing if he had accompanied the material with more critical comments regarding its scientific and statistic value.

The *second section* of the chapter on dictionaries for foreign-language learners is dedicated to bilingual dictionaries. It is generally accepted that most bilingual dictionaries are of a lower quality than their monolingual counterparts. A careful reading of Welker's text does not change this negative opinion,

but it is to his merit that he once again is able to present the material in such a way that it becomes interesting and worth reading. He first deals with bilingual dictionaries in general and then with bilingual dictionaries for foreign-language learners. In the first part, he discusses and provides examples of some specific problems related to bilingual lexicography such as *typology*, *equivalents*, *metalanguage*, *meaning differentiation* as well as *lemma selection*, *separation of homographs*, and *differentiation and arrangement of senses*.

In the second part, he starts with a "short history" of what he calls "preoccupations about bilingual dictionaries for foreign-languages learners" and then discusses "the characteristics of good dictionaries" of this type. Here he applies the same methodology as in some of the previous chapters and refers to a large number of scholars such as the Russian Ščerba, the Pole Tomaszczyk, the Brazilians Gomes de Matos, Amaral and Schmitz, the North Americans Iannucci and Lindstrom, the Arab Al-Kasimi, the Germans Hausmann, Herbst, Werner and Zöphen, the Spaniard Alvar Ezquerro, the Frenchman Dubois, the Briton Atkins, the Italian Celotti, the Belgian Humblé, etc. In this way — apart from discussing some of the most important bilingual dictionaries for foreign-languages learners — he succeeds in showing the development of theoretical thinking about this type of dictionaries as well as the most relevant ideas presented by scholars from various traditions and language communities. The result of this guided tour through the world of bilingual learners' lexicography is that the reader, even when disagreeing with many of the ideas expressed, becomes inspired to formulate new ideas that may eventually lead to an improved lexicographical practice.

In this manner, the whole chapter on monolingual and bilingual dictionaries for foreign-language learners turns into a must-read for anybody who wishes to work theoretically and practically within this field.

6. Native-language pedagogical dictionaries

Welker himself writes that the major part of the existing theoretical literature about pedagogical lexicography is dedicated to dictionaries for learners of foreign languages. This disparity may be the reason why the previous chapter on such dictionaries is almost three times as long as the one on dictionaries for native-language learners. But this does not mean that it is less interesting. Welker divides these dictionaries into *children's dictionaries*, *school dictionaries* and *desk dictionaries* (*college dictionaries*).

The *section about children's dictionaries* begins with a quotation from Hausmann (1990: 1365) which is reproduced here because of its relevance for the most important characteristics of this type of dictionary:

- a) The layout is especially clear. Space is not saved. The letters are bigger than in general dictionaries. Colours are generally used. The dictionaries have frequently a big format.
- b) All lemmata, or a considerable part of them, are illustrated.
- c) There are no definitions; or when there are, they are not conventional.

d) Narrative texts (lexicographic story-telling) substitute the traditional macrostructure. e) There is no information about the lemma, or when it is provided, it is only very little. f) Abbreviations are not used. g) Exercises are given. h) The macrostructure is very selective, never with more than 5,000 lemmata. Generally it is between 200 and 2,000 lemmata. i) In most cases, the lemmata refer to concrete things. j) The foreseen users are children below 10 years.

According to Hausmann, there are two types of children's dictionaries, the *visual dictionaries* organized thematically and the *alphabetic dictionaries*. Hausmann also notes that there is no clear division between children's dictionaries and school dictionaries and that their titles often do not reveal their real content. This claim is supported by Welker who subsequently refers to the observations made by various scholars — mainly Brazilians — who describe a number of children's dictionaries designed for children up to 14 years of age, with much more lemmata (up to 30 000), and with some other characteristics different from the ones Hausmann emphasized.

The problem here seems to be that some of these scholars confuse commercial titles and scientific categories, that some of the dictionaries treated are not adequate for the expected user group, and that *children's dictionary* and *school dictionary* are imprecise and misleading terms. In most countries, children start school between the ages of 5 and 7, are considered children at least up to the ages of 12 or 14, and continue in school up to the age of 15 or 16. This means that they, for a long period, are *school children*. Consequently, if a school dictionary is defined as a dictionary conceived to be used by the pupils in school, most school dictionaries are at the same time "children's dictionaries". The real distinction that should be made is therefore between *preschool dictionaries* and *school dictionaries* whereas the latter should be graduated and subdivided into various categories according to the pupils' mental and linguistic development and their growing knowledge about the world. In this respect, the characteristics described by Hausmann in the above quotation should therefore be viewed as the necessary characteristics of both preschool dictionaries (especially the visual ones) and school dictionaries for children (pupils) below the age of 10. Although Welker himself does not discuss these complex questions related to typology, his treatment of the field does nevertheless allow observant readers to draw conclusions of their own.

In the *section about school dictionaries for mother-tongue learners*, Welker (p. 302) starts by noting that — surprisingly — such dictionaries are not found in all countries and that, even in countries where they have been published, the following distinction should be made:

It is necessary to distinguish the genuine dictionaries from the ones that are only smaller versions — with less lemmata, senses and information — of general dictionaries. "Genuine" means "really pedagogical" and especially prepared for the pupils of primary or secondary school.

He then, once again, provides a large number of reflections on English, French, German, Italian, Spanish and Portuguese school dictionaries as well as some of

the results of the investigation conducted in Brazil into the use of Portuguese school dictionaries. The main topics discussed in this section is the type of school dictionary, the number of lemmata, the formulation of the definitions, and the type of additional lexicographical data assigned to the lemmata. Welker starts by referring to Hausmann (1989) who considers the fathers of the modern pedagogical dictionary both for foreign-language and mother-tongue learners to be Harold Palmer, Michael West and especially Edward Thorndike who published the *Junior Dictionary* (1935), the *Senior Dictionary* (1941) and the *Beginning Dictionary* (1945).

However, this opinion is somehow opposed by Pruvost (2001) who counts 64 French school dictionaries since the publication of Pierre Larousse's *Nouveau Dictionnaire de la Langue Française* in 1856, i.e. several decades before Palmer, West and Thorndike marked the era of English pedagogical lexicography. Pruvost (2001: 74) also makes an important distinction between "reductionist dictionaristics" and "heuristic lexicography". In the first case, the school dictionary is produced by means of a simple reduction of a "more important dictionary" for adult users, whereas "heuristic lexicography" presupposes an original work with a macro- and microstructure "linguistically and psychologically adapted to the learner". In this respect, Welker himself (p. 303) writes:

Hence, it is clear that it is necessary to write the definitions or explications very carefully in order to improve the comprehension, not only for foreign-language learners but also for native-language pupils.

The whole section about school dictionaries for mother-tongue learners contains much interesting and useful information, not least about dictionaries produced in Brazil which are discussed with many details allowing the reader to form an idea of the impressive work done by practical and theoretical lexicographers in Brazil during the last few years.

In the *section about desk dictionaries* (which are called "usage dictionaries" in Portuguese), Welker initially discusses the two terms *desk dictionary* and *college dictionary* (also called *collegiate dictionary*) and observes that some lexicographers distinguish between them, whereas others consider them to be synonymous. Welker (p. 332), who does not have this terminological problem in Portuguese, observes that these dictionaries are not always regarded as pedagogical dictionaries, "because they are not specifically designed for language learners", but he himself considers them as such, describing them as follows:

Generally speaking, desk dictionaries are monolingual dictionaries designed for native speakers, with a minor macrostructure than the big general monolingual dictionaries and different from these in the sense that they present certain important components for text production.

Subsequently, Welker first analyzes some English, Spanish and Italian desk dictionaries which he calls "atypical" because of their large size, and then some other French, Spanish and Italian dictionaries which are more "normal". The

section ends with a detailed study of a number of Portuguese desk dictionaries in which Welker mainly refers to his own previous publications and comments on this dictionary type. This section as well as the rest of the chapter on dictionaries for mother-tongue learners which is just as inspiring as the previous chapter on dictionaries for foreign-language learners, is another must-read for those who work theoretically and practically with pedagogical dictionaries.

7. Special pedagogical dictionaries

In this chapter, Welker deals with the types of pedagogical dictionaries which are considered "special", because they are not "normal", i.e. they "are neither normal monolingual pedagogical dictionaries nor normal bilingual pedagogical dictionaries" (p. 357). He discusses ten different types of these special pedagogical dictionaries:

- (a) Hybrid bilingual dictionaries
- (b) Multifunctional monolingual dictionaries for foreign learners
- (c) Encyclopedic dictionaries for foreign learners
- (d) Onomasiological and analogical dictionaries
- (e) Visual dictionaries
- (f) Valency dictionaries
- (g) Dictionaries of collocations
- (h) Dictionaries of idioms
- (i) Dictionaries of false friends
- (j) Pedagogical dictionaries of specialized languages

Of course, this list could easily be extended, e.g. with special dictionaries of common learner errors, as suggested by Frankenberg-Garcia (2010). However, the chapter shows above all the great variety of pedagogical dictionaries commercialized on a market which has to come up with ever new products in order to survive. Some of these dictionaries provide information which the user may find in other dictionaries, and frequently of a higher quality, although some of them are interesting as supplements to already existing dictionaries, especially in the case of printed dictionaries. As usual, Welker takes the reader on a well-guided tour, this time in a new region of the world of pedagogical dictionaries, calling on a number of scholars to express their opinions about the various lexicographical solutions furnished by these dictionaries. As it is impossible to reproduce all these opinions and discuss the various types of special pedagogical dictionaries in the framework of this article, only two of them will be commented on here. The first is the *encyclopedic dictionary for foreign learners* which, among other matters, provides cultural and encyclopedic information needed with regard to the foreign-language learning process. In this respect, Welker quotes Zgusta (1989: 3) for the following opinion:

However, since language is embedded in culture, cultural data are important to the learner not only for steering his linguistic behaviour but frequently for choosing the correct lexical equivalent. Such cultural information can be understood in a broad way, so that it can pertain to political and administrative realities of the country or countries whose language is being learned, and so on. Undoubtedly, a good part of this information is of encyclopedic character; be this as it may, it belongs to what the learner has to learn.

If Zgusta is right, and he most certainly is, and if cultural data are needed to steer the learner's linguistic behaviour and ability to choose the correct words, then the question is why lexicographical data of this type should be included in a separate dictionary type instead of the "normal" dictionary for foreign-language learners. Welker himself does not answer this question but refers to Stark (1999) who distinguishes between "compulsory" and "optional" cultural and encyclopedic information of which the first type is necessary to define the lemma appropriately. In this sense, the "normal" dictionary for foreign-language learners should furnish the compulsory information, while the optional information could be provided by special encyclopedic dictionaries.

The second type of special dictionary to be commented on here is the pedagogical dictionary of specialized languages. In this respect, Welker mainly goes into details with Jean Binon and Serge Verlinde's *Dictionnaire d'apprentissage du français des affaires* (DAFA) on which he has also commented in a previous chapter. The very existence of this and other similar dictionaries reveals a fact which is often forgotten in theoretical discussions, i.e. that language learning is learning not only of the so-called general purpose language, but also of special languages, both in the mother tongue and in foreign languages. Although this type of specialized dictionary as well as the theoretical literature accompanying it is still relatively limited in comparison to other products of pedagogical lexicography, it nevertheless constitutes an area of growing interest, a fact proved by two theoretical books on this topic, i.e. Fuertes-Olivera and Arribas-Baño (2008) and Fuertes-Olivera (2010), published since the appearance of Welker's book. Welker should therefore be praised for introducing this type of dictionary to a lexicographical community who frequently ignores it in spite of its importance for a large number of language learners.

8. Electronic dictionaries

In a tree-diagram on page 27, provided after a short discussion on the typology of pedagogical dictionaries, Welker makes a primary-level distinction between printed and electronic dictionaries which are then subdivided into monolingual and bilingual dictionaries, CD-ROM and Internet dictionaries, etc. Another possible typology could consist of a primary distinction between pedagogical dictionaries for learners of the mother tongue, foreign languages and

science, and a secondary distinction between dictionaries for general purpose language learning and special purpose language learning, etc. The first of these typologies is based on formal criteria (media and number of languages) whereas the second is bound up with the functions of the dictionaries. None of these typologies is better or worse than the other. It all depends on the purpose of the typology. However, what is important is that electronic dictionaries are not fundamentally different from printed dictionaries in terms of their content and functions. As to this, Welker (p. 415) starts the chapter on electronic dictionaries with an observation which he has already previously made in the first chapter:

All types of dictionaries can exist in electronic format. The characteristics which existing electronic dictionaries have in common with printed dictionaries will not be presented here.

Instead, Welker focuses on some of the new aspects of electronic dictionaries. Initially it should be noted that any book on this topic will be more or less obsolete when it finally appears owing to the rapid development and continuous flow of new products. This is also, up to a certain degree, the case with Welker's book. According to him, electronic dictionaries can be divided into:

- (a) Dictionaries used in processing of national language;
- (b) CD-ROM dictionaries;
- (c) Online dictionaries (accessible via Internet); and
- (d) Dictionaries on hand-held computers (pocket dictionaries).

Welker observes that the first type mentioned is not regarded by all researchers as an electronic dictionary. However, there is at least one type of electronic dictionary which he does not mention, although it had been available in some countries, e.g. South Africa (but maybe not Brazil), when he wrote the book, i.e. dictionaries on mobile phones. Welker then refers to a large number of electronic pedagogical dictionaries as well as to the opinions expressed by other scholars. As to the advantages of electronic dictionaries, he himself comments (p. 419):

The major advantage of online dictionaries — and electronic dictionaries in general — is the search facilities.

This judgement may be valid for 99 percent of all electronic dictionaries. It is perfectly true that the use of electronic media optimizes the search facilities but there is another advantage which until now has only been partially explored by a few dictionaries, i.e. the possibility to adapt the dictionary to the specific needs of each user in each user situation (cf. Tarp 2011). Welker himself (p. 420) calls for more "inclusion of multimedia", referring to Lemberg (2001):

The ideal would be to be able to access the pronunciation of the lexemes, the images of the referents, maybe the sound produced by the verbs *crepitar* (crunch)

and *ranger* (squeak), and even videos showing actions like *torcer* (wrestle) and *driblar* (dribble).

It is evident that these possibilities should be exploited. But the real ideal would be to go beyond the present practice and develop lexicographical pedagogical e-tools which can be adapted to the needs of each student. In such a world, the user, when consulting the dictionary, would not obtain all the lexicographical data assigned to a specific lemma, but only the data needed in each situation. This would, for instance, also make it possible to take the data from each of the special pedagogical dictionaries mentioned in the previous chapter and put them into one and the same data base from which it could be extracted in an individualized form by means of an advanced interface allowing interaction between the student and the dictionary.

However, such a world will not become reality only by observing present-day lexicographical practice. It is necessary not only to learn, but also to "unlearn" and leave behind bad habits as observed by Gouws (2011). What is needed is an advanced lexicographical theory capable of transforming the discipline and creating a new generation of dictionaries that fully adapt to the new technologies. Such a theory will have to be the result of the concentrated intellectual efforts of many lexicographers who, among other matters, will have to base their research on a meticulous study of the practice hitherto. In this sense, neither the chapter about electronic dictionaries nor the rest of Welker's book presents any solution. However, it greatly facilitates the work to be done by providing easy access to a large amount of material produced by generations of practical and theoretical lexicographers.

9. Conclusions

The *Panorama geral da lexicografia pedagógica* consists of a total of 519 pages, of which the last 65 pages are back matter containing three indexes of authors, dictionaries, and key terms, all of them with references to the specific pages, as well as a list of all the books, articles and other theoretical contributions to which the author refers. Together with the detailed table of contents (an English translation is available at http://vsites.unb.br/il/let/welker/LP_contents (accessed July 6, 2010)), this variety of access options makes it easy not only to read the book from cover to cover, but also to consult it and return to it whenever necessary.

Welker's book is written to satisfy the specific needs of a Brazilian audience desiring lexicographical knowledge but it deserves to be read by many more scholars within the international lexicographical community. The book may be used as a reference work which, in a condensed form, provides easy access to many different opinions and ideas with which it is impossible always to agree because of their contradictory and sometimes even completely opposing character, but which is nevertheless a source of inspiration to anybody

doing theoretical and practical work within the sphere of pedagogical lexicography — and even beyond.

Although it would have been useful to have more theoretical input from Welker himself, it is exactly the descriptive method he applies that gives the book this character of a much needed reference work or encyclopedia on pedagogical lexicography. Of course, it is always possible to find dictionaries not mentioned, authors not quoted, and ideas not reflected, as indicated by Frankenberg-Garcia (2010) in her review of the book, but it cannot be denied that Welker is an extraordinarily diligent and meticulous author who has produced a comprehensive work far more inspiring than, for instance, Dolezal and McCreary (1999).

Welker's book is highly recommendable and deserves, as already stated, to be read also by lexicographers who are not conversant with Portuguese text reception at an academic level. If translated into English, it would probably become a "best-seller" within a lexicographical community that is becoming ever more aware that theoretical lexicography cannot be constrained to the British and European tradition. Hopefully, a publishing house will come forward to accept the challenge.

Literature

- Basel, Barbara and Louise Banks (Eds.).** 2006. *Illustrated Dictionary of Natural Sciences and Technology Today*. Cape Town: Maskew Miller Longman.
- Beattie, Nicholas.** 1973. Teaching Dictionary Use. *Modern Language* 54(4): 161-168.
- Binon, Jean and Serge Verlinde.** 1999. *Dictionnaire d'apprentissage du français des affaires*. Paris: Didier.
- Bogaards, Paul.** 1991. Dictionnaires pédagogiques et apprentissage du vocabulaire. *Cahiers de Lexicologie* 59(2): 93-107.
- Bosman, D.B., I.W. van der Merwe and A.S.V. Barnes.** 2002. *Tweetalige Skoolwoordeboek Afrikaans-Engels/Engels-Afrikaans*. Cape Town: Pharos.
- Cowie, Anthony P.** 1996. The 'Dizionario Scolastico': A Learner's Dictionary for Native Speakers. *International Journal of Lexicography* 9(2): 118-131.
- Cowie, Anthony P.** 1999. *English Dictionaries for Foreign Learners — A History*. Oxford: Clarendon Press.
- Dolezal, Fredric T. and Don R. McCreary.** 1999. *Pedagogical Lexicography Today. A Critical Bibliography on Learners' Dictionaries with Special Emphasis on Language Learners and Dictionary Users*. Tübingen: Niemeyer.
- Dziemianko, Anna.** 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Tübingen: Max Niemeyer.
- Frankenberg-Garcia, Ana.** 2010. [Review of] Herbert Andreas Welker. *Panorama Geral da Lexicografia Pedagógica*. Brasília: Thesaurus. 2006. *International Journal of Lexicography* 23(2): 233-236.
- Fuertes-Olivera, Pedro A. and Ascensión Arribas-Baño.** 2008. *Pedagogical Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins.

- Fuertes-Olivera, Pedro A. (Ed.).** 2010. *Specialised Dictionaries for Learners*. Berlin/New York: Walter de Gruyter.
- Gouws, Rufus H.** To appear. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). To appear. *e-Lexicography. The Construction of Internet Dictionaries*.
- Gu, Peter Yongqi.** 2003. Vocabulary Learning in a Second Language: Person, Task, Context and Strategies. *TESL-EJ* 7-2 [online]. <http://tesl-ej.org/ej26/a4.html> (Accessed July 6, 2010).
- Hartmann, R.R.K. and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hartmann-Petersen, Preben, Gus Gerrans and Rasmus Hartmann-Petersen.** 2001. *Sasol Science & Technology Resource. A Book Resource for Learners, Students and Trainees in Science and Technology*. Observatory/Claremont: Infosource.
- Hausmann, Franz Josef.** 1989. Kleine Weltgeschichte der Metalexikographie. Wiegand, Herbert Ernst (Ed.). 1989. *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 75-109. Tübingen: Niemeyer.
- Hausmann, Franz Josef.** 1990. Das Kinderwörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 1365-1368. Berlin/New York: Walter de Gruyter.
- Herbst, Thomas.** 1985. Von Fehlern, die vermeidbar wären. Ein weiteres Argument für mehr Wörterbucharbeit in Englischunterricht. *Bielefelder Beiträge zur Sprachlehrforschung* 14(1–2): 139-165.
- Hernández, Humberto Hernández.** 1998. La lexicografía didáctica del español: aspectos históricos y críticos. Fuentes Morán, María Teresa and Reinhold Werner (Eds.). 1998. *Lexicografías ibero-románicas: problemas, propuestos y proyectos*: 49-79. Madrid/Frankfurt am Main: Iberoamericana and Vervuert.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. New York: Scribner Press.
- Lemberg, Ingrid.** 2001. Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. Lemberg, Ingrid, Bernhard Schröder and Angelika Storrer (Eds.). 2001. *Chancen und Perspektiven computergestützter Lexikographie*: 71-91. Tübingen: Niemeyer.
- Lew, Robert.** 2007. [Review of] Herbert Andreas Welker: *O Uso de dicionários: Panorama geral das pesquisas empíricas*. Brasília: Thesaurus. 2006. *International Journal of Lexicography* 20(4): 401–403.
- Martínez de Sousa, José.** 1995. *Diccionario de lexicografía práctica*. Barcelona: Bibliograf.
- Mathews, Mitford M.** 1955. The Freshman and his Dictionary. *College Composition and Communication* 6: 187-190.
- Pálfi, Loránd-Levente and Sven Tarp.** 2009. Die Lernerlexikographie in Skandinavien. Entwicklung, Kritik und Vorschläge. *Lexicographica. International Annual for Lexicography* 25: 135-154.
- Pruvost, Jean.** 2001. Les dictionnaires d'apprentissage monolingues de la langue française (1856–1999): Problèmes et méthodes. Pruvost, Jean (Ed.). 2001. *Les dictionnaires de langue française. Dictionnaires d'apprentissage, Dictionnaires spécialisés de la langue, Dictionnaires de spécialité*: 67-95. Paris: Honoré Champion.

- Rundell, Michael.** 1998. Recent Trends in English Pedagogical Lexicography. *International Journal of Lexicography* 11(4): 315-342.
- Stark, Martin.** 1999. *Encyclopedic Learners' Dictionaries: A Study of their Design Features from the User Perspective*. Tübingen: Niemeyer.
- Sutton, Kathleen (Ed.).** 2007. *Multilingual Science Dictionary for South African Schools*. Cape Town: Longman.
- Tarp, Sven.** 2008a. Første oversigtsværk over leksikografiske brugerundersøgelser. [Review of] Herbert Andreas Welker: *O Uso de dicionários: Panorama geral das pesquisas empíricas*. Brasília: Thesaurus 2006. *LexicoNordica* 15: 343-345.
- Tarp, Sven.** 2008b. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, Sven.** 2009. Reflections on Lexicographic User Research. *Lexikos* 19: 275-296.
- Tarp, Sven.** 2010. Reflections on the Academic Status of Lexicography. *Lexikos* 20: 450-465.
- Tarp, Sven.** To appear. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, Pedro A. and Henning Bergenholtz (Eds.). To appear. *e-Lxicography. The Construction of Internet Dictionaries*.
- Welker, Herbert Andreas.** 2004. *Dicionários. Uma pequena introdução à lexicografia*. Brasília: Thesaurus.
- Welker, Herbert Andreas.** 2006. *O Uso de dicionários: Panorama geral das pesquisas empíricas*. Brasília: Thesaurus.
- Welker, Herbert Andreas.** 2008. *Panorama geral da lexicografia pedagógica*. Brasília: Thesaurus Editora.
- Wiegand, Herbert Ernst.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 13-30. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst (Ed.).** 1998. *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von Langenscheidts Großwörterbuch Deutsch als Fremdsprache*. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst (Ed.).** 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuch Deutsch als Fremdsprache*. Tübingen: Max Niemeyer.
- Xatara, Claudia and Philippe Humblé (Eds.).** 2006. *Tradução e lexicografia pedagógica*. Cadernos de tradução XVIII. Florianópolis: Universidade Federal de Santa Catarina.
- Zgusta, Ladislav.** 1989. Idle Thoughts of an Idle Fellow; or, Vaticinations on the Learners' Dictionary. Tickoo, Makhan L. (Ed.). 1989. *Learner's Dictionaries: The State of the Art*: 1-9. Singapore: SEAMO Regional Language Centre.
- Zöfgen, Ekkehard.** 1994. *Lernerwörterbuch in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. Tübingen: Max Niemeyer.

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Der-tiende Deel*: R. 2009, xxi + 541 pp. ISBN-13: 978-0-9584971-7-6 (plastiekband), ISBN-13: 978-0-9584971-8-3 (leerband). Stellenbosch: Buro van die WAT. Prys: R290 (plastiekband) / R590 (leerband).

Hierdie resensie is die beskouing van 'n taalpraktisyn wat geen leksikograaf is nie. Die uitgangspunt is dus nie leksikografiese teorie nie en die werk word ook nie aan leksikografiese teorie beoordeel nie. Dit is die beskouing van 'n leksikografiese leek, maar 'n origens min of meer ingeligte gebruiker.

Wat 'n genotvolle ervaring was dit nie om in hierdie woordeboek te lees nie! Want dit was wat ek inderdaad gedoen het om hierdie resensie te skryf — nou nie elke artikel op elke bladsy van **1R** af tot by **rywiël** nie, maar darem heelwat — afgesien daarvan dat ek Deel XIII al 'n paar maande lank daaglik werk toe en terug ry (**1ry I 4**) om te gebruik.

Ek kan sonder huiwering sê dat dit uiters interessante leesstof is. 'n Mens sien nie net woorde wat vir jou onbekend is en hul betekenis en gebruik nie, maar jy verbaas jou telkens weer vir al die betekenisonderskeidings wat daar aan bekende woorde sit en vir die bondige, knap wyse waarop die redaksie dit definieer. Bowenal getuig Deel XIII weer daarvan dat die WAT as geheel 'n kultuurskat van die hoogste orde is. Die lees daarvan verplaas 'n mens opnuut na die Afrikaanse leefwêreld in sy volle omvang en ryke (**1ryk 2 a, 4, 5**) verskeidenheid, oud en nuut.

Terselfdertyd gee dit plek-plek egter ook 'n blik op hoe Engels Dietse vorme reeds verdring het en steeds verdring — soos "vir die res" (s.v. **res I**, UITDR. GESPECIALISEERD; hoekom dit 'n gespesialiseerde uitdrukking is en nie 'n idiomatiese uitdrukking nie, is nie vir my duidelik nie) en "so reg soos reën" (s.v. **1reg II** UITDR. B IDIOMATIES), of "'n rollende klip vergaar geen mos nie" (s.v. **rol I** UITDR. IDIOMATIES). In plaas van sulke kwalik verbloemde Engelse uitdrukkings kan daar eerder plek afgestaan word aan 'n eg Afrikaanse (hoewel miskien nie baie verfynd nie!) uitdrukking soos "jou agterent/gat in rat kry" saam met die ander uitdrukkings onder **1rat I** UITDR. B IDIOMATIES. Die WAT se redaksionele beleid vereis egter dat dit onder **agterent** en/of **gat** behandel moes gewees het en nou is dit heeltemal vir die gebruikers verlore.

Gepraat van 'n kultuurskat: Op p. i noem die hoofredakteur lakonies die Buro se finansiële posisie, en op die teenoorstaande bladsy word 'n lang lys donateurs genoem. Die Buro se ankerdonateurs en elke ander individu of organisasie wat daar genoem word, verdien werklik die lof en dank van elke Afrikaanssprekende vir wie die taal kosbaar is.

"Afrikaanse" maatskappye — by name dié wat hulle bestaan uit Afrikaans maak, soos uitgewershuse e.d. — en organisasies wat in die breë Afrikaanse gemeenskap werksaam is en veral Afrikaanse mense as lede of kliënte het, behoort as 'n prioriteitsaak die Buro van die WAT finansiël te ondersteun sodat die redaksie hom met voldoende personeel aan sy regte (**1reg II 5**) taak kan wy, naamlik om die woordeskat van Afrikaans so omvattend moontlik op

te teken en leksikografies te bewerk. Om die WAT as omvattende woordeboek van Afrikaans te voltooi en by te werk of te hersien — veel het immers met die taal en sy woordeskat gebeur sedert die eerste dele gepubliseer is — is seker die belangrikste monument wat ooit vir Afrikaans en al die sprekers daarvan opgerig kan word.

Die voorwerk noem (op pp. i-iv) ook 'n merkwaardige lys vakgebiede en vakkundige medewerkers wat die redaksie van deskundige raad dien (¹**raad 1**; hier, terloops, moes die oorspronklik korrekte uitdrukking *van raad dien* minstens naas die neologistiese *van raad bedien* as kollokasie vermeld gewees het, al kom eg. in 'n sitaat voor) by die hantering van veral vakwoorde. Dit bied nie net 'n uitstekende beeld van die reikwydte (**reikwydte 2**) van Afrikaans en die terreine waarop Afrikaans hom handhaaf of laat geld nie, maar dui ook weer op hoe bereid Afrikaanse mense gewoonlik is om hulle kennis sonder vergoeding in "die saak" terug te ploeg. Dié vakkundige medewerkers is donateurs van 'n ander aard, en verdien eweneens dank en eervolle vermelding. En enig-een wat sou wou beweer dat Afrikaans nie 'n volledig ontwikkelde akademiese, wetenskaps- en kultuurtaal is nie, kan maar net na hierdie lys (en natuurlik die woordeboek self) kyk.

Die boek is, soos gewoonlik, netjies en duidelik uitgelê: alles is in helder, baie leesbare fonte gedruk. Van p. vi tot p. xix word gebruikstoeligtig gegee. Die eerste twee bladsye hiervan bevat 'n benoemde skematiese voorstelling van die inskrywings, wat besonder nuttig is om die soms nogal ingewikkelde verduidelikings op die volgende paar bladsye te illustreer. Wat vir my egter effens vreemd was in dié skema, is die keuse van **reguithed** om 'n onverklaarde *samestelling* te illustreer — veral omdat dit onder die lemma **reguit** verskyn, sou ek dit eerder as 'n afleiding beskou. Of hanteer die redaksie samestellings en afleidings as dieselfde (sien **samestellings** links bo, p. xi)? Vir enigiemand wat die inhoud van die woordeboek werklik ten volle wil verstaan en tap, is die gebruikstoeligtig van beslissende belang, al kos dit soms kophou om alles uit te pluus.

Die gebruikstoeligtig bevat wel 'n paar glipse, soos byvoorbeeld dat by die woordverklaring van **artikel** (p. viii) "ingelui" gebruik word in plaas van die hier korrekte "ingelei", of die verlore spasie tussen "35" en "mm" by *35mm-kamera* onder 2.3 op p. xi. Afdeling 2.2 op p. xi sê ook dat loopkoppe boaan elke bladsy "in vet hoofletters" aangegee word, maar vir my lyk die loopkoppe nie na vetdruk nie.

Ek vind dit ook moeilik om te begryp waarom daar op skynbaar lukrake wyse afgewyk word van die gevestigde punktuasiekonvensie wat punte binne of buite slothakies betref. Op byvoorbeeld p. xvii word op een na (by **kaart: rooi kaart**) al die punte buite die slothakies geplaas al is dit telkens 'n selfstandige volsin tussen die hakies. Die hoofletters waarmee hierdie parentetiese sinne telkens begin word, laat my verder vermoed dat die punt nie tot afsluiting van die hele paragraaf dien nie, maar slegs van die betrokke sin.

Nog iets rakende konvensie: Die konvensie betreffende taksonomiese nomenklatuur waarmee ek reeds meer as drie dekades bekend is (en by implika-

sie in die *Afrikaanse woordelys en spelreëls* [AWS] opgeneem is — sien reël 9.8), is dat generieke en spesifieke name gekursiveer word maar dat hoër klassifikasies romein bly. Tog gebruik Deel XIII ook vir die hoër klassifikasies, soos "familie", deurgaans kursief (kyk byvoorbeeld "(fam. *Mimosaceae*)" onder **rooihaak** op p. vii). Moontlik is dit die navolging van 'n styl wat in die vroeëre dele begin is, maar as 'n mens in ag neem hoeveel die jonger dele wel van vroeëres verskil, is dit nie vir my 'n geldige rede vir afwyking van die konvensie nie.

Wat die liggaam van die teks betref, maak die feit dat die betekenis- en woordsoortonderskeidings, soos in die laaste klompie dele, telkens op 'n nuwe reël (**reël 1**) begin met die toepaslike nommer in vetdruk, die gebruik van die papieruitgawe baie makliker aangesien die onderafdelings van elke artikel in duidelik sigbare eenhede uiteenval. Die enigste kritiek wat ek oor hierdie aspek van Deel XIII het, is dat indien daar onder "Uitdrukings" sowel gespesialiseerde as idiomatiese uitdrukings verstrekkend word, die afdeling "B IDIOMATIES" soms baie moeilik raakgesien word. Dié probleem sou oorkom kon word deur 'n reëlspasie (**reëlspasie 1**) telkens daarvoor in te voeg.

In taalkundige kringe word seker algemeen aanvaar dat woordeboeke nie as taalvoorskriftelik beskou moet word nie, met ander woorde dat woordeboeke nie preskriptief is nie maar deskriptief. Prof. Rufus Gouws skryf byvoorbeeld onlangs: "'n Woordeboek skryf nie voor hoe taal gebruik moet word nie, maar beskryf eerder hoe taal werklik gebruik word" ("Help taal duidelik* word", *Beeld*, 7 April 2010, p. 19). Almiskie, maar die deursneegebruiker ervaar 'n woordeboek enigsins anders — miskien nie altyd as heeltemal voorskriftelik nie, maar wel as meer of minder normatief. 'n Woord of uitdrukking is aanneemlik of korrek "want dit staan in die woordeboek". Regsprekers raadpleeg woordeboeke om die betekenis(se) van woorde of uitdrukkinge vir doeleindes van hul uitsprake (vonnisse) te bepaal, mense gebruik woordeboeke om die korrekte spelling van woorde te bepaal of om vas te stel of 'n sekere woord of uitdrukking deel van die betrokke taal uitmaak (en dus gebruik "mag" word), of om uit te vind presies hoe 'n uitdrukking of idioom gebruik (moet) word, en so meer. Hoe ook al, vir gewone woordeboekgebruikers het 'n woordeboek skynbaar baie groter voorskriftelike of normatiewe waarde as vir die leksikograaf.

Om dié rede behoort die leksikograaf m.i. baie versigtig met die materiaal om te gaan sodat die woordeboek nie net aan die heersende leksikografiese teorie voldoen nie, maar ook die sienings en sensitiwiteite van en gebruik deur die "gewone gebruiker" verreken. Onlangs het 'n uitgewer my byvoorbeeld gebel om my standpunt te verneem oor die opname van sekere woorde in 'n sekere verklarende Afrikaanse woordeboek. Die rede vir die oproep was dat 'n studenterradiostasie die spot gedryf het met die betrokke woordeboek weens die opname van woorde soos *cool*, *show*, *stunning* en *stupid*. Die feit dat sulke woorde op groot skaal in gesproke Afrikaans gebruik word, was nie vir die studente ter sake nie. Vir hulle was dit gewoon nie Afrikaans nie, en daarmee basta.

Die WAT moet dus besonder goed nadink oor die opname van woorde soos *rally*, *rampaartie*, *rent*, *rep*, *roly-poly*, *rondshunt*, *rondorder* en dergelike, wat almal in Deel XIII voorkom, soos trouens ook die sinonieme *moewie* en *movie* (s.v. **rolprent**), selfs al word dit as "geselstaal" geëtiketteer. Dat die woorde dikwels voorkom, selfs in skriftelike taal in hoofstroom- of standaardtaalpublikasies, kan seker nie betwyfel word nie (sommige sitate bewys dit immers), maar die meeste gebruikers voel dit instinktief aan as "onsuiwer" of "nie Afrikaans nie" en dus as iets wat nie in 'n woordeboek soos die WAT tuis hoort nie. In aansluiting hierby betwyfel ek dus die geldigheid (vir opname) van 'n sitaat soos "Buurvrouens deel alles, ook roly-poly's wat vir boyfriends bedoel is" (s.v. **roly-poly**) of "Hy was 'n rep, toe retrench hulle hom" (s.v. **rep**), selfs al word die betrokke lemmas geëtiketteer as "(<Eng.)" of "(*geselstaal*)".

Teenoor bogenoemde soort gevalle staan natuurlik leen- of volksetimologiese vertalings soos **raaigras** (p. 6). Dié vorm kom reeds in die oudste landboubron wat ek tot my beskikking het (Du Plessis, S.J. 1950. *Voorlopige Woordelys in Landbou en Bosbou*. Stellenbosch: Pro Ecclesia), naas *roggras* voor. In 'n vorige bedeling het ek jare lank met die werk van die destydse landboudepartemente in die staatsdiens te doen gehad, en ek kan my nie herinner dat ek een keer *roggras* teëgekom het nie — *raaigras* was die standaardterm. Deel XIII bevestig dit deurdat **roggras** (p. 333) nie as sodanig gedefinieer word nie maar die gebruiker terugverwys na **raaigras**, waar die artikel die definisie bevat. Die woord **raaigras** is dus dermate verafrikaans en ingeburger dat bitter min mense waarskynlik sal besef dat dit regstreeks aan Engels ontleen is, wat nie van *rep*, *rondshunt* e.d. gesê kan word nie.

En so gepraat van ontlenings aan Engels: onder **rekord** word onder idiomatiese uitdrukkinge "**vir die rekord** (*n.d. Eng.* for the record)" ingeskryf, maar 'n veelvoorkomende uitdrukking soos "prokureur van rekord", wat insgelyks "na die Engels" is, verskyn nie, want dit moes volgens die WAT se redaksionele beleid reeds onder **prokureur** behandel gewees het. (Terloops, terwyl ek met die regte doenig is: om by die afkorting **R.** te beweer dat dit "ongewoon" en "meer dikwels" **regt.** is, betwyfel ek werklik, want daar is nie 'n Afrikaanse hofverslag of regshandboek e.d. waarin "R" ná die betrokke regter se naam nie gebruik word nie.)

Daar is ook ander woorde of uitdrukkinge wat ek nie kon vind nie, soos *regdeur* (slegs as onverklaarde samestelling vermeld); *rekenaarhardeware* (*rekenaarsagteware* is wel daar); *responsoriaal*, *responsories* en *responsorium*; *rondtrek* en *rondtrekkend* (waarvan daar in die e-WAT ewe veel gevalle voorkom as *rondreisend*, wat wel opgeneem is [s.v. **rondreis**]); *Roomse gevaar* en so meer. Ek kan my nie voorstel dat hierdie woorde en uitdrukkinge so selde in die geëkserppeerde bronne voorkom dat dit nie opname regverdig nie. Onder invloed van Engels "throughout" kom *regdeur* byvoorbeeld deesdae so dikwels voor dat 'n mens feitlik nie meer die "ou" *deur die hele* ... sien nie. Ek is ook geen voorstander van die gebruik van "hardeware" en "sagteware" in rekenaarkonteks nie, maar dat dit op groot skaal gebruik word, is onmiskenbaar. Bowendien, as *reke-*

naarsagteware wel opgeneem word, is daar kwalik regverdiging vir die weglating van *rekenaarhardeware*.

Die weglating van items soos bogenoemde is seker geen doodsonde nie. By my ontstaan die vraag egter waarom sulke "leemtes" in die WAT voorkom. Dit is bekend dat onder die hoofredakteurskap van dr. Dirk van Schalkwyk die Buro 'n beleidsverandering met betrekking tot die afhandeling van die woordeboek aanvaar het en dat dele nou in 'n bepaalde tempo (en van beperkter omvang?) afgehandel moet word. Die gebruiker kan dus met reg wonder of sulke leemtes (en dié wat ander resensente reeds uitgewys het) per ongeluk voorkom, of die betrokke woorde of uitdrukkings nie aan die norme vir opname voldoen nie, of ruimtebeperkings opsetlike weglatings meebring, en of dit die gevolg van die versnellingsbeleid is. Indien laasgenoemde, is dit bitter jammer, want dan loop die WAT gevaar om nie meer werklik 'n omvattende Afrikaanse woordeboek te wees nie.

Indien ruimtebeperkings 'n oorweging is, is daar werklik maniere om ruimte te bespaar sodat meer lemmas behandel kan word. Die opvallendste hiervan is m.i. die fonetiese transkripsies van die lemma-uitspraak. Ek het dit allermens teen die aangee van die uitspraak — dit is 'n baie verrykende toevoeging tot die jongste paar dele — en in verreweg die meeste gevalle is die aanbod van uitspraak bondig en hanteerbaar. Ek dink egter dat daar plek-plek 'n bietjie oordadig omgegaan word met kleiner uitspraakverskille. Myns insiens behoort die fonetiese transkripsies beperk te word tot hoogstens drie of vier variasies van wat die standaarduitspraak is of behoort te wees. Die lys transkripsies by 'n lemma soos **radiomedium** (dit beslaan amper sewe reëls), **reduksiepotensiaal** en **reduksionisme** (nege reëls elk) of **revolusionisme** (elf reëls) is so oorweldigend dat geen gebruiker daardeur gaan worstel nie. Hier word elke denkbare uitspraakvariasie skynbaar verreken. Daarenteen word by lemmas met 'n [œy]-klank, soos in **reguit**, die veel voorkomende [æi]-uitspraak ("regyt") tereg nie verreken nie. (Vreemd genoeg word by lemmas met byvoorbeeld **rug-** die uitspraak met 'n schwa [rɔx] egter wel aangegee.)

Terloops, ek dink as 'n mens in die dae toe daar nog Afrikaans in die Weermag was, vir die militêre mense sou sê "die [rɔvɔi:] word gespeel" (s.v. **reveille**), sou hulle óf nie verstaan het nie, óf gelag het. Ek het dit nog nooit in ons weermagkonteks anders gehoor as iets soos "rewellie" nie ...

Ek meen ook dat die hantering van wisselvorme soms 'n bietjie rojaal met ruimte omgaan. Om byvoorbeeld twee afsonderlike artikels vir **rekenaargebaseer** en **rekenaargebaseerd** en drie afsonderlike artikels vir **rekenaarfoendi**, **rekenaarfoendie** en **rekenaarfundi** te hê (natuurlik met die nodige kruisverwysings), is m.i. onnodig (en plek-plek selfs verwarrend).

Insgelyks sou 'n paar reëls ruimte bespaar kon word indien lemmas wat in die meervoudsvorm opgeneem word, onder die enkelvoudsvorm gehanteer word. So 'n geval is **revolusies**. In die betekenis wat hier bedoel word, word dit inderdaad altyd in die meervoudsvorm gebruik, maar dit is m.i. niks anders as die meervoud van **revolusie 2** nie, soos die verwysing na **omwenteling 2 a ii** trouens bewys.

Ek het hierbo gesê die WAT verplaas die leser na die Afrikaanse leefwêreld, oud en nuut. Wat 'n genot is dit nie om opnuut of vir die eerste maal kennis te maak nie met woorde soos:

ramas o.m. 'driftig, luidrugtig en vernielsugtig te werk gaan';
rapallie 'gepeupel';
remonte 'vars perd';
rens 'effens suur'
ringeloor 'op onwaardige wyse dwing of behandel';
rinnewiet 'iemand wat goed verinnuweer of verniel';
rofkas — word mure nog ooit gerofkas?;
rondas 'ronde skild';
rondwaar 'verward of geheimsinnig soos 'n spook ronddwaal'; en
roof/roffie/rower — het ons hedendaagse weermag nog rowe/rofies?

Dalk is dit iets wat die Buro van die WAT as 'n reklameveldtog kan aanpak: om skole of koerante uit te daag om nuwe lewe aan ou woorde te gee in plaas van om die altyd oorheersende Engels slaafs na te volg.

Deel XIII gee egter nie net aan ou woorde aandag nie. Die wêreld waarin ons tans lewe, word eweneens verreken. Daarvan getuig die opname van 'n hele konstellasië terme wat met rekenaars verband hou — byvoorbeeld 'n stuk of ses bladsye vol samestellings met **rekenaar**- as eerste lid — asook lemmas soos **rap** (daardie "musiek"-styl), **rassetamboer**, **roofkyker** en **roofluisteraar**. Dit gee soms ook 'n idee van hoe vinnig ons (tegnologiese) wêreld verander — dink maar: net 'n paar jaar gelede was 'n **ratslot** (p. 75) iets wat elke motoreienaar geken het, maar wat vir die jonges van vandag seker amper antiek is.

Ek het hierbo opgemerk dat die redaksie goed moet nadink oor die opname van lemmas of sitate wat wesenlik Engels of anglisisties is, en ook dat gewone gebruikers 'n woordeboek in meerder of mindere mate as normatief beskou. Gedagtig hieraan bly dit vir my altyd 'n vraag in watter mate die redaksie "taalfoute" — dit wil sê grammatikale, sintaktiese, idiomatiese of spelingsafwykings van die heersende norme — in sitate moet regstel (buiten natuurlik sulke afwykings wat aan die ouderdom van die sitaatbron toegeskryf kan word). Ek self vind dat ek, veral met die beskikbaarheid van die e-WAT, allerlei elektroniese soektogte doen om te bepaal of 'n bepaalde woord of konstruksie wel of nie "Afrikaans" is. Op grond van wat ek dan vind of nie vind nie, besluit ek oor die aanvaarbaarheid al dan nie van die betrokke vorm.

So byvoorbeeld het Deel XIII my oortuig dat "ten regte" (**1reg I** UITDR. A GESPECIALISEERD) 'n aanneemlike uitdrukking is (en dat dit nie, soos ek destyds geleer is, "tereg" hóéf te wees nie). Maar wat gemaak met 'n sitaat soos "Navorsing dui aan dat die meeste gewapende rowe gepleeg word deur persone van tussen 16 jaar oud en in die vroeë twintigs" (s.v. **2roof II 1 b**) — "van tussen ... en in ..."? Dit is tog sintakties nie pluis nie. Ook "Die elektriese ruitopdraaiers, wat al vier syvensters beheer, is standaard toerusting" (s.v. **ruitopdraaier**). Tans word "standaard" nog nie as 'n attributiewe adjektief aanvaar nie, en die

skryfwyse behoort dus "standaardtoerusting" te wees. Verwant hieraan is die skryfwyses van die redaksie self wat nie met die heersende spelreëls ooreenstem nie, soos die lemma **Rehoboth-Afrikaans**, wat volgens reël 14.14 van die negende en die tiende uitgawe van die AWS eintlik "Rehobothafrikaans" behoort te wees, en "Duitswes-Afrika" (s.v. **rebellie 2**), wat immers "Duits-Wes-Afrika" behoort te wees (sien AWS, 10de uitgawe, reël 12.31).

Ek het ook enkele setfoutjies opgemerk, soos die woordafbreking by die gespesialiseerde uitdrukking **regs om** onder **regs II** (REG-SOM behoort m.i. REGS-OM te wees) en 'n punt wat skynbaar weggeval het tussen "komposita" en "Selde" in die tweede reël van **1regs-**.

Die kritiek wat ek in hierdie beskouing uitgespreek het, val egter amper in die kategorie van vittery. Dit doen geensins afbreuk aan die waarde en nut van hierdie deel van die WAT nie. Ek kan my nie voorstel hoe ek my beroep na behore sou kon beoefen sonder die WAT nie, en vir enige taalwerker is dié deel dus 'n belangrike en noodsaaklike toevoeging tot sy of haar beroepsgereedskap. Al waaroor ek bitter spyt is, is dat ek my beroepsloopbaan sal moet afsluit sonder om die vreugde en gebruik van 'n volledige WAT te gesmaak het.

J.D. McLachlan
Departement Taaldiens
Universiteit van Suid-Afrika
Pretoria
Suid-Afrika
(*tommcl@mweb.co.za*)

G.-M. de Schryver (Editor). *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks.* 2010, viii + 376 pp. ISBN 978-9970-101-01-6 (Hardback). Linguistics Series. Kampala: Menha Publishers. Price: €59.95.

A Way with Words is a Festschrift volume to honour the 70th birthday of Patrick Hanks. The volume is attractively printed on thick, good quality paper, hardbound in cloth and supplied with a cover sleeve (which, I admit in shame, I have a habit of throwing away, but this one has an adorable portrait of Patrick which can be enjoyed quite independently of the book). The title on the cover is rendered in golden letters, perhaps a hint at the level of contribution the dedicatee has made to the study and description of meaning.

The book is divided into three parts, each representing a major domain of Patrick Hanks' activities: lexical theory with a bit of philosophy of language ("Theoretical Aspects and Background"), computational and corpus-driven lexicology ("Computing Lexical Relations"), and the analysis and synthesis stages of the lexicographic process ("Lexical Analysis and Dictionary Writing"). From the inspiring introductory chapter by Gilles-Maurice de Schryver, we learn that the editor has attempted to reflect, within each of the three sections, the path of development of the Festschrift's ideas. This is a brilliant concept, and the execution is just as outstanding.

In fact, the introduction ("Getting to the Bottom of How Language Works") is an important part of the book, and it makes fascinating reading. In it, we are offered an insight into the origins of the book, and then the focus moves to Patrick Hanks himself, outlining his contribution to lexical theory (including, in some more detail, the latest *Theory of Norms and Exploitations*, the topic of a forthcoming book), corpus linguistics, and lexicography — the milestones of his professional career — as well as his publications, whose complete listing immediately follows the introductory chapter. The final section of the introduction explains the structure and content of the book, which leaves the reader well primed for the remaining nineteen contributions.

To save some space, I will not give here a full title for every single chapter, especially as these are conveniently available on the publisher's website.¹ Instead, let me say a sentence or two about what each piece has to offer, all being original, previously unpublished articles (with the possible exception of Green's contribution on French *argot* dictionaries, which is an extended version of his recent publication).

As the opening chapter in Part I of the book, we already find a veritable collector's item. This is a heretofore unpublished essay by the late John Sinclair (a close collaborator of Patrick Hanks), wherein he uses the verb *sever* to propose a descriptive approach that gives proper balance to aspects of form and meaning, all the while stressing the point that the word alone is not a unit of language. The paper is unfinished but no less worth reading. Following it is a contribution by Yorick Wilks, who makes a case for a Preference Semantics approach in the area of natural text processing. Next, James Pustejovsky and

Anna Rumshisky deploy the artillery of the Generative Lexicon to tackle the difficult issue of sense extension in verbs. Igor Mel'čuk, the giant of lexical semantics, develops the notion of the Government Pattern in the setting of the Explanatory Combinatorial Dictionary (and, more broadly, Meaning-Text Theory). This part of the book is concluded with a piece by the logician David Wiggins, who in effect probes the philosophical soundness of the lexicographer's daily work.

Part II starts with an article by Ken Church. Although titled "More is More", it happens to be the shortest of the chapters. I, for one, did not find much new in it, although some readers may actually enjoy its informal and digressive style. In contrast, I found the next chapter by Greg Grefenstette highly original and fascinating. Using the texts from the Web as indexed by Exalead,² he ambitiously attempts to estimate the number of two-word multi-word units in English, arriving at 200 million of (what he terms) concepts. In the next contribution, David and Louise Guthrie assess the potential of English adjectives to give away the semantic category of the head noun they modify. In their computational experiment, they predict that not all adjectives will have equally useful discriminating power, and they consider several criteria here: co-occurrence with semantically unambiguous nouns, the information-theoretic measure of entropy (a higher entropy would indicate a less distinct pattern of distribution), and the frequency of reoccurrence in the training set. Under the more favourable conditions, the accuracy of semantic category assignment is on the order of 70%, which is a promising result. German light verb constructions are the topic of an article by Alexander Geyken. His findings point to an advantage of gigaword corpora: the latter turn out to be capable of identifying most useful verb-noun expressions, but the same cannot be said of a balanced 100-million-token corpus. In the chapter to follow, Karel Pala and Pavel Rychlý take a critical look at the output produced by the Sketch Engine. Taking a close look at the word sketch for the Czech verb *vidět* 'see', they identify errors in the sketch output, concluding that the problems are mainly due to tagging errors and imperfections in the sketch grammar, and offering suggestions as to how these could be improved. The two final chapters of Part II tie in directly with the *Pattern Dictionary of English Verbs*, a project in progress coordinated by Patrick Hanks. The first is a report by Silvie Cinková, Martin Holub, and Lenka Smejkalová of an attempt to develop experimental evaluation procedures with which to test the consistency of entries already in the *Pattern Dictionary of English Verbs*, especially with respect to the assignment of Semantic Type labels to noun collocates of verbs. In the second, Elisabetta Jezek and Francesca Frontini discuss their plans to extend the Corpus Pattern Analysis technique (which underlies Pattern Dictionaries), a proposal for linking back the verb patterns to their corpus context by annotating the latter with pattern tags. The paper outlines a plan for an implementation of such a *Patternbank* for Italian, and examines the potential benefits of the extension.

Part III of the book opens with a contribution by Rosamund Moon, who undertakes a finely detailed corpus study of the English phrase *spring to mind*,

including a classification of the various uses into textual functions and speech acts. Sue Atkins then presents the new exciting Database of Analysed Texts of English, or DANTE,³ which will soon provide the breadth and depth of lexical information for English on a level not seen before. Next, Adam Kilgarriff (also one of the chief architects of DANTE) and Pavel Rychlý revisit the enticing scenario of tick-box lexicography (which here gets the more formal, if less catchy, name of semi-automatic dictionary drafting), which may well be the most successful project so far to take the drudgery out of dictionary-making. Following this comes a thought-provoking position paper by Paul Bogaards, wherein he argues against the idea of an independent (meta)lexicographic theory. Mirosław Bańko then retells the story of how his fascination with the COBUILD dictionary led to the creation of a similar dictionary for the Polish language. The freelance lexicographer specializing in slang, Jonathon Green, presents a concise historical overview of French dictionaries of slang. Part III of the volume, and the volume itself, concludes with Michael Rundell's dissection of the concept of *elegance* in lexicography. *Elegance* is an elusive attribute, and even if the author does not fully succeed in providing the definitive genus and differentia of *elegance*, his piece offers a solid dose of lexicographic common sense, tastefully adorned with amusing anecdote. What an elegant way to wrap up this unique volume!

To conclude, in *A Way with Words* Gilles-Maurice de Schryver has managed to put together a highly coherent and worthwhile collection of articles written by a stellar constellation of authors, and artfully structured in a way that traces the achievements of Patrick Hanks, to whom the volume is dedicated. Needless to say, the level of editorship is as high as you would expect from this editor. This is a must-read for any serious student of word meaning, word use, corpus analysis and lexicography.

Endnotes

1. See <http://www.menhapublishers.com/products.html#tablecontents>.
2. See <http://www.exalead.com>, which also provides a little-known, but perhaps the most linguistically-aware search engine available (<http://www.exalead.com/search/>).
3. See <http://www.webdante.com>, which, at the time of this writing, allows searches for headwords in the M–R letter range, but access will likely soon be extended to the complete database.

Reference

Hanks, P. Forthcoming. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.

Robert Lew
Department of Lexicology and Lexicography
School of English
Adam Mickiewicz University
Poznań, Poland
rlew@amu.edu.pl

G.-M. de Schryver (Editor). *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. 2010, lviii + 582 pp. ISBN-13 978 0 19 576554 0 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R104.95.

The *Oxford Bilingual School Dictionary: Zulu and English* (henceforth OZSD) is the latest addition to the bidirectional English–isiZulu bilingual dictionary market and is based on the same successful and prize-winning formula used for the *Oxford Northern Sotho School Dictionary* (ONSD) published in 2007.

One could state at the outset that OZSD is an excellent dictionary — as in the case of ONSD it brings together what could simply be called all of the *best lexicographic practices* for isiZulu. The spirit of *learn isiZulu and English* runs throughout the entire work, whether consulting the central texts or reading the guidelines and Study Section — the compilers left no stone unturned to guide users in finding the information they are looking for.

The front matter of the OZSD gives a table of contents, a user-friendly explanation of the dictionary features and an introduction. The Study Section located between the isiZulu–English and the English–isiZulu components contains the mini-grammar as well as guidance to dictionary activities, writing of e-mails and letters, spelling and pronunciation, etc. The back matter consists of a Reference Section on animals, fruit and vegetables, the human body, etc. These plates and tables successfully bring together items decontextualised as an inevitable result of alphabetical ordering in dictionaries. Reading the Study Section is important in order to decode the information when looking up words in the OZSD.

Dictionary articles are elaborate and give information on, among others, frequency of use, part of speech, sense distinction, translation equivalents, examples of usage in source and target language — in a very user-friendly way. Consider the articles of the verb *-fundela* and the noun *isihambi* in (1) and (2) as typical examples.

- (1) **-fundela** *verb + applicative* **C** -FUNDA
1 ■ learn; study ♦ UVusi ufuna **ukufundela** ubudokotela noma ubumeli. *Vusi wants to study either medicine or law.* 2 ■ read for ♦ Bakhuthazwe ukujwayela **ukufundela** ukuzijabulisa. *It is necessary that they be encouraged to get used to reading for enjoyment.*
- (2) **isihambi** *noun* 7/8 (*pl. izihambi*) **C** -HAMB
■ traveller; visitor ♦ Akekho osaziyo **isihambi** lesi. *Nobody knows this traveller.*

Text or shade(d) boxes is an innovative feature for isiZulu dictionaries and enhance the quality of the treatment given in the OZSD. They give guidance in respect of nouns used as adverbs, grammar (especially past tense forms), combinations, spelling, word division, etc.

The presumed coverage of English and isiZulu by the OZSD (p. xii), i.e. 82% and 71%, is impressive.

The compilers decided to use English as the metalanguage for both components of the dictionary. Using isiZulu as metalanguage could also be considered as an option in future revisions. This decision is questionable — especially in a school dictionary where all other aspects and presentations are punctiliously done on an equal basis for the two languages.

The most innovative feature in OZSD is the lemmatisation of the full form of nouns instead of stem forms as has traditionally been done in isiZulu dictionaries, i.e. the abandoning of the stem tradition for the lemmatisation of nouns. Even frequent locative forms that are derived from nouns using the so-called *e-/o-...-ini* 'locativization strategy' have been listed (p. S18). Full forms of adjectives, relatives, adverbs and pronouns can also be directly looked up.

Compare OZSD's lemmatisation of *umntwana* and *abantwana* with the traditional way, e.g. as in Doke and Vilakazi's *Zulu-English Dictionary* (ZED) in (3) and (4) respectively.

(3) OZSD (word lemmatisation)

umntwana ... (*pl.* abantwana)
... child
abantwana ... *pl. noun* 1/2 See sg. UMNTWANA

(4) ZED (stem lemmatisation)

-ntwana (umntwana ... abantwana)
baby, small child ...

Although there has been some debate in the past decade regarding the applicability of lemmatising nouns according to their full forms, i.e. prefix plus noun, no lexicographer or lexicographic unit to my knowledge thus far dared to break this almost sacred tradition of stem lemmatisation for nouns. Stem lemmatisation is even regarded as the scientifically correct way of lemmatisation and this presumed superiority over word lemmatisation influenced dictionary compilation for African languages to such an extent that it has even been used in some dictionaries for disjunctively written languages such as Sepedi, Sesotho and Setswana where it was much more questionable than in dictionaries for conjunctively written languages such as isiZulu.

Many lexicographers have come to the erroneous conclusion that only the stem tradition is linguistically justified. Ziervogel ... for example, claims that it is scientifically sound, and Ziervogel and Mokgokong ... state categorically that it is the only scientific method. (Van Wyk 1995: 84)

A first typical argument against the lemmatisation of the full forms of nouns is that the alphabetical stretches into which the noun class prefixes fall will be overcrowded. This means that the vowel sections would be very large, e.g. words beginning with *izi-* (classes 8 and 10), *um-* (classes 1 and 3), *aba-* (class 2),

etc. will result in very long alphabetical stretches in the dictionary. In OZSD it is indeed the case that *i-* takes up 62 pages, representing 23.5%, i.e. almost a quarter of the dictionary, *u-* 40 pages, etc. but users are unlikely to find it disturbing in any way.

The second argument is that identification of the lemma entails the selection of a section from the complex orthographic word anyway — therefore why not then also cut the noun prefixes, see discussion below.

What is important, however, is that OZSD, by lemmatising full forms of nouns, manages to avoid a number of lemmatisation problems resulting from certain phonological processes such as aspiration (cf. Van Wyk 1995). For example, stem identification is very problematic in cases such as *intaba*, *intombi*, *inkosi*, *inkabi*. The user would not know that the stem form of *intaba* is *-ntaba* but for *intombi* it is *-thombi*, for *inkosi* *-khosi* and for *inkabi* *-nkabi* in order to look it up in a traditional isiZulu dictionary. In OZSD the problem is avoided by the lemmatisation of the full nominal forms *intaba*, *intombi*, *inkosi* and *inkabi*.

If viewed from the angle of the user, it is true that when, for example, the teacher instructs him/her to write an essay on *umsebenzi* he/she would be best served by an isiZulu dictionary giving the natural (full) form *umsebenzi* as lemma. The problem arises when the need for look-up entails complex words in the learner's prescribed literature. The core of the problem is that verbs as well as nouns often occur with huge clusters of circumfixes: prefixes, suffixes, conjunctives, etc. Consider the arbitrary selection from the hundreds of verbal and nominal forms found in the corpus pertaining to the verb *-sebenza* 'work' or one of its verbal or nominal derivations. The OZSD offers numerous articles, for the lemmas *abasebenzi*, *ekusebenzeni*, *ekusebenziseni*, *ekusetshenzisweni*, *emisebenzini*, *emsebenzini*, *imisebenzi*, *isisebenzi*, *izisebenzi*, *-sebenzela*, *-sebenzisa*, *-sebenzisana*, *-setshenziswa*, *-setshenzwa*, *ukusebenza*, *ukusetshenziswa*, *umsebenzi*, but the challenge for the user is to identify the appropriate lemmas from the complex orthographic words in (5). (Note, however, that the frequency of occurrence of many of these forms is low.)

- (5) *abasebenzelayo*, *bayasebenza*, *belusebenzisa*, *bengasawasebenzisi*, *emsebenzini*, *komsebenzi*, *kusebenze*, *lomsebenzi*, *lusebenzisa*, *msebenzi*, *nasemsebenzini*, *nawumsebenzi*, *nemisebenzi*, *okuzisebenzela*, *olisebenzisayo*, *olungasenamsebenzi*, *ozokuwasebenzisa*, *sengimsebenzele*, *sesizozisebenzisa*, *ubuzakungisebenzela*, *zomsebenzi*

Although lemmatisation of the full forms of isiZulu nouns goes a long way in easing the problem of stem identification, the problem of lemma identification is by no means solved and will never be in paper isiZulu dictionaries. The need to identify even full forms of isiZulu nouns for look-up remains problematic in some instances especially for the inexperienced user. Consider also *umuntu* 'human being' in this regard. Once again, if the user is confronted with the noun in isolation as in the case of *umsebenzi* above, there is no problem and it is much more natural and easy not to have to remove the *umu-* in order to look up *-ntu* in OZSD, in contrast to traditional dictionaries. However, in cases

where the user is confronted with complicated orthographic forms containing/involving *umuntu* such as *njengomuntu*, *ngingumuntu*, *lowomuntu* in (6), the target user could find it difficult to identify the lemma *umuntu*. From a frequency point of view (frequencies indicated between brackets), it has to be noted that the noun itself, i.e. *umuntu* in this case, is overwhelmingly more frequent in an isiZulu corpus than the other orthographic forms. With a little experience, and using the user guide of OZSD, the user will probably be able to restore the missing pre-prefix in order to look up *-muntu* under *umuntu*.

- (6) *umuntu* <19479>, *muntu* <4314>, *lomuntu* <1754>, *ngumuntu* <1239>, *nomuntu* <1058>, *kumuntu* <795>, *ungumuntu* <768>, *njengomuntu* <746>, *yomuntu* <695>, *komuntu* <467>, *ngomuntu* <337>, *somuntu* <318>, *okomuntu* <307>, *womuntu* <297>, *ngingumuntu* <228>, *engumuntu* <211>, *namuntu* <181>, *lowomuntu* <145>, *zomuntu* <140>, *wayengumuntu* <119>, *bomuntu* <117>, *omuntu* <87>, *kungumuntu* <78>, *kunomuntu* <77>, *kwakungumuntu* <75>, *zamuntu* <71>, *wumuntu* <61>, *nalomuntu* <60>, *umunt* <52>, *ingumuntu* <46>, *kulomuntu* <44>

So, although the many virtues of lemmatisation of full nominal forms is not disputed, it does not mean that a golden solution has been found to noun or noun stem identification. OZSD guidelines are, however, honest and clear on this issue in trying to give guidance in dealings with 'however's' and 'furthermores', e.g. p. S17, in cases such as *-muntu* in (6).

A potential problem lies within the lemmatisation of infinitives. OZSD correctly lemmatised full forms of a number of frequently used infinitive nouns such as *ukudla* 'food', *ukuhamba* 'departure' and *ukukhuluma* 'speaking' in the alphabetical stretch *uku-* and treated them appropriately for their nominal meanings. These infinitive nouns, however, stand in contrast with the infinitive verbs *ukudla* 'to eat', *ukuhamba* 'to walk/go' and *ukukhuluma* 'to speak' in isiZulu. *Ukudla*, *ukuhamba* and *ukukhuluma* have therefore also correctly been lemmatised in OZSD under their stem forms *-dla*, *-hamba* and *-khuluma* with applicable treatment for their verbal meanings. However, for looking up verbs in isiZulu dictionaries users become used to also looking up infinitive verbs under their stem forms. Thus removing the infinitive prefix *uku-* by default prior to look-up, could result in a problem where the user does not consider the possibility that he/she should also check under *uku-* for the possible existence of an infinitive noun with the same stem. I believe that a cross-reference is required from the articles of the verb stems to the full nouns in such cases. Consider the following examples. Say for instance the user is confronted with the infinitive noun *ukujula* 'depth' but looks it up as for an infinitive verb, i.e. under *-jula*, he/she will find the meaning 'consider carefully' which is not applicable in this case. A cross-reference from *-jula* to *ukujula* would have solved the problem. Inserting such cross-references would of course require additional space in the dictionary. The alternative should be to warn users in the guidelines to check for possible nominal forms when looking up infinitives. If the same line of argumentation is followed for *ukuthi*, the user might not find the

meaning of the most frequently used word in isiZulu, the conjunctive *ukuthi* 'so that' in the absence of a cross-reference from *-thi*¹ ... *verb* ... *say* ... to *ukuthi*.

Lemmatising both singular and plural forms of nouns is especially recommended for learners' dictionaries. This, however, comes at a huge price in terms of redundancy of space taken up by lemmatising the other member of the pair, usually the plural forms. Once again the compilers of the OZSD took the best option, that is, lemmatising the plural forms, and instead of treating them, they supplied a cross-reference to the singular form as in (7).

Lemmatising plural forms with cross-referencing to the singular forms results in an overuse of the mediostructure (cross-referencing) as lexicographic device, rendering sections as in (7) that consist entirely of cross-references in OZSD.

- (7) **izimpendulo** * *pl. noun 9/10* See sg. IMPENDULO
izimpi * *pl. noun 9/10* See sg. IMPI
izimpiko *pl. noun 11/10* See sg. UPHIKO
izimpilo *pl. noun 9/10* See sg. IMPILO
izimpisi *pl. noun 9/10* See sg. IMPISI
izimpondo *pl. noun 11/10* See sg. UPHONDO
izimpophoma *pl. noun 9/10* See sg. IMPOPHOMA
izimpukane *pl. noun 9/10* See sg. IMPUKANE
izimu *noun 5/6 (pl. amazimu)* ■ cannibal
♦ *Izimu* balinikeza imbiza evuzayo. *They gave the cannibal a pot with a hole in it.*
izimvu *pl. noun 9/10* See sg. IMVU
- izindlela** *** *pl. noun 9/10* See sg. INDLELA
izindlovu *pl. noun 9/10* See sg. INDLOVU
izindlu ** *pl. noun 9/10* See sg. INDLU
izindondo *pl. noun 9/10* See sg. INDONDO
izindonga *pl. noun 11/10* See sg. UDONGA
izinduku *pl. noun 9/10* See sg. INDUKU
izinduna ** *pl. noun 9/10* See sg. INDUNA
izindwangu *pl. noun 9/10* See sg. INDWANGU
izindwani *pl. noun 11/10 (sg. udwani)*
■ grass blades
◇ **-dla izindwani** ■ be overflowing; be in flood ♦ Ngathi lapho ngifika eKwiti, ngawufica udla izindwani. *When I came to the Kwiti River, I found it to be in flood.*

This, however, is defensible and OZSD in my view even understates the value of its lemmatisation of plural forms of nouns in saying that "the only information you will find there is a cross-reference to the singular form" (p. S15). In (5), reconstructing *imisebenzi* from *nemisebenzi* and actually finding *imisebenzi* as a lemma in the dictionary is firstly a confirmation of successful lemma identification, secondly frequency of use information, thirdly noun class information and finally a cross-reference to the singular where full treatment is given. These are all important bits of information given with lemmatised plural forms of nouns in OZSD.

As in the case of ONSD, it is not clear what the exact title for reference purposes of OZSD should be: the outside cover refers to *Oxford, ISIZULU-ISINGISI, ENGLISH-ZULU, Isichazamazwi Sesikole, School Dictionary* and the first title page to *Isichazamazwi Sesikole Esinezilimi Ezimbili, ISIZULU NESINGISI, Esishicilelwe abakwa-Oxford, Oxford Bilingual School Dictionary, ZULU AND ENGLISH*. Sticking to this complicated naming template was probably done to be on a par with the other dictionaries in the series but the practical value is questionable. Colour, font and layout do, however, contribute to the readability of the title, see

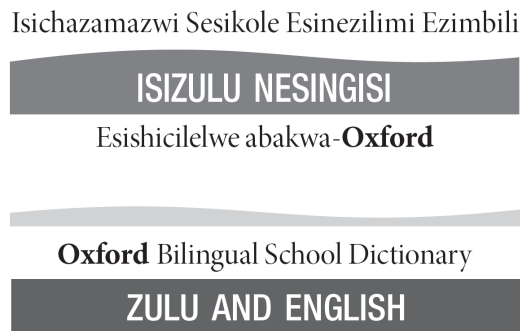


Figure 1: Title page of the *Oxford Bilingual School Dictionary: Zulu and English*

Affordability is a limiting factor for African language dictionaries. Bidirectional dictionaries bridging English with an African language in South Africa are currently caught up in a triangulation of number of lemmas versus exhaustiveness of treatment versus price, cf. Prinsloo (2009). This simply means that 500 to 600 pages are the default limit within which the compiler can operate as prescribed by the publishers. In principle, these limitations leave the compiler with two basic options: the inclusion of a large number (e.g., 20 000 to 30 000) of lemmas with limited (e.g., 1 to 2 lines double column) treatment, or a limited number (e.g., 10 000) of lemmas with more exhaustive (e.g., 5 to 7 line) treatment. The market price is normally limited to R100 per dictionary. The OZSD provides extended/exhaustive treatment but consequently the number of lemmas is limited to 5 000 in the Zulu to English section and 5 000 lemmas in the English to Zulu section.

Conclusion

In the past decade, a number of studies have been undertaken to establish best practices in terms of lemmatisation, balancing of alphabetical stretches, combating inconsistencies, compilation of corpus-driven dictionaries for African languages, etc. The problems inherent in lemmatisation are real. These studies were performed against the background of the user-perspective. School dictionaries must, by definition, be easy to use. It can be concluded that publication of the OZSD represents a new era for isiZulu–English lexicography because the latest insights, lexicographic tools, an isiZulu corpus and a state-of-the-art dictionary writing system have been utilised. The OZSD succeeds in its aims to offer support in the key areas of helping learners choose the right translation, giving frequently used translations, showing how words are really used, the inclusion of new words from across the curriculum as well as the incorporation of useful additional components such as a mini-grammar.

References

- ONSD = **De Schryver, G.-M. (Ed.)**. 2007. *Oxford Bilingual School Dictionary: Northern Sotho and English*. First Edition. Cape Town: Oxford University Press Southern Africa.
- OZSD = **De Schryver, G.-M. (Ed.)**. 2010. *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. Cape Town: Oxford University Press Southern Africa.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the *Oxford Northern Sotho School Dictionary*. *International Journal of Lexicography* 22(2): 151-178.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- ZED = **Doke, C.M. and B.W. Vilakazi**. 1948. *Zulu-English Dictionary*. First Edition. Johannesburg: Witwatersrand University Press.

D.J. Prinsloo
Department of African Languages
University of Pretoria
Pretoria
South Africa
(danie.prinsloo@up.ac.za)

Jean Nicolas De Surmont. *Chanson: son histoire et sa famille dans les dictionnaires de langue française.* 2010, IX + 248 pp. ISBN 978-3-484-52353-1 (Hardback). Berlin/New York: Walter de Gruyter. Price: €49.95

The fundamental aim of the study of the *chansonnier* vocabulary in French is to highlight the existence of a literary genre, namely the *chanson*. The author states that the *chanson* genre, as referred to in the Middle Ages, is characteristically French in view of the prestige of the French language and culture over centuries and the study thus comprises an analysis of the term *chanson* and its lexical family from its origins in old French to today's modern French.

The author begins with a theoretical and methodological approach to the study of the terminology of various song genres, with reference to works and dictionaries on lexicography, emphasizing that the term *chanson* assumes different sung or written forms. The aim is to carry out a semasiological and lexical study based on a linguistic corpus as well as a study of *chansonnier* phenomena, using works on the history of French and Quebec literature as the main field of reference. In the case of Quebec culture, the study is limited to the historiography of the last forty years. The range of the corpus is discussed and the corpus chosen to represent elements relating to the song phenomenon is defined as being part of a group of essays on the *chanson*, alphabetic dictionaries, glossaries or vocabularies relating to terminology or specialised dictionaries of sayings or idiomatic expressions. Extensive use is made of dictionaries in French from the 15th century onwards in a thorough investigation of the term *chanson*.

The author then proceeds to explain the way in which the corpus is analysed. This metalinguistic corpus consists of printed and electronic elements as well as a series of essays alluding directly or indirectly to the definition of song genres. Discussions centre on the type of vocabulary used and the metalinguistic terms describing it and the range of vocabulary to be studied is situated somewhere between general and specialised language. Operating concepts are suggested within the song vocabulary and are defined as a *phénomène chansonnier*. This phenomenon encompasses all the practices, manifestations or traditions relating to the song-object as well as research and publications on the song. This comprehensive term refers to all social, historical and historiographical facets of *chanson*. This first theoretical section concludes with a brief discussion of various types of poetry, such as sung or vocal poetry (which the author prefers to the term oral poetry) as well as the difference between a signed *chanson* (of unknown origin) and a song from oral tradition. The concept of lexical engineering is introduced as part of an operating vocabulary to be used in analysing different aspects of word groupings, idiomatic expressions and collocations in the ensuing chapter.

Having established a theoretical background, the next section deals with an analysis of the lexical family of the song with its double linear components of music and text. At this point the author introduces the concept of referential

contamination which revolves around the double origin of the *chanson* and refers back to 16th century poets like Du Bellay who were conscious of the interplay between music and poetry. The troubadours and trouvères are cited as examples. Reference is also made to modern artists like Brassens who sings poems by Villon and Hugo which gives rise to the question what is the dominant linearity in song: the text or the music? There are further references to the 19th century French poets who debated the effects of poetry and musicality. In an effort to define the nature of the song, the author postulates that the term *chanson* can only be justified by the context. According to the numerous sources referred to, *chanson* can be a volume of poetry, melodies, operettas, poems, choirs, romances set to music as well as a wide range of different musical works. A further differentiation is made between oral and sung poetry. In a synchronic analysis of the lexicographical and encyclopaedic corpus the same definitions are suggested where the changing nature of song activity is emphasized from French texts drawn from the 17th century to the present. Various types of *chansons* are analysed in the light of musical dictionaries including the continuous interrelationship between text and music. The author concludes that both text and music are equally significant components of the song.

The study continues with a diachronic analysis of the *chanson* family and begins with a history of the meanings of the term from Latin through to the present day, taking into consideration numerous literary traditions. As a result of the lack of grammatical rules and the influence of orality, a large number of lexical forms appear, such as *chanchon*, *cançon* and *chanson* in the 14th century. The author then establishes a semantic cartography of the *chanson* such as *chanson* and *canço* as used in the Middle Ages, together with variations attested in lexicographical sources. Similar analyses are carried out during the period from the 12th to the 20th century. There follows a detailed and interesting exploration of the lexical field in terms of *chanson*. The list begins with diminutive forms such as *chansonnette* which itself also covers a wide semantic spectrum, including pejorative forms such as can be seen in Boileau's *Art poétique*. *Chan-telette* and the noun form *chansonnier* are also derived from *chanson*, meaning a volume of secular songs in the Middle Ages as well as a writer or composer of songs as referred to by various 18th and 19th century authors. Writers and composers of satirical songs and spiritual songs are alluded to (Calvin refers to *chansons spirituelles*). From the 16th to the 19th century *chansonnier* also takes on a political connotation. Modern dictionaries refer to variations in the use of the term *chansonnier* and in Quebec *chansonnier* assumes divergent forms from traditional French interpretations. Feminised derivatives like *chansonnière*, verb forms *chansonner*, *se chansonner*, and noun forms *chansonneur* are discussed.

An analysis follows of the meaning of the *chanson* lexeme according to a metalinguistic corpus. The large number of dictionaries consulted result in a dynamic vision of the language. Both the strict meaning and the figurative sense are presented, with *chanson*, for example, taking on the meaning of a harmonious noise. Examples are also given from the 14th century onwards of

pejorative or familiar meanings, metaphorical variations as well as related terms resulting from metonymy. For example lexical co-existents of *chanson* such as *air*, *romance*, and *rengaine* are discussed. Similarly, idiomatic units are analysed following a definition of an idiomatic expression as a group of words used in a fixed sense which does not correspond with the full sense of the individual words and is governed by semantic rules. This includes proverbs. A detailed enumeration follows, drawn from a linguistic and metalinguistic corpus, of expressions derived from *chanson* such as *J'ai en ça pour une chanson* with direct English equivalent of 'I got it for a song.' Again, examples are drawn from the 14th century to the present, including Quebec sources. The author then discusses collocations, defined as such in 1951, where it is the expression that carries the meaning and not the individual words. Examples are given such as *chanson populaire* and *chanson de toile*. Similar occurrences are referred to in English and Hungarian where the extension of a word in one language is sometimes wider than in another such as the deeper meaning of *chanson* in comparison to song. Over the centuries numerous phrases have been used by French authors to describe the song phenomenon such as *chanson galante*, *chanson bachique* and more specifically in the oral tradition, *chanson d'eau* and *chanson enfantine* amongst many others. The syntagms derived as above are clearly descriptive of specific activities whether drinking, dancing, heroic exploits (the well-known *chanson de geste* as found in Chrétien de Troyes as well as variants *chanson de bonne geste*), singing ballads, playing the lute, reference to a hero (*La chanson de Roland*) or royalty (*chanson royale*).

In conclusion, in the author's words, linguistic exchanges, borrowings and calques have enriched the semantic field of *chanson*. The need is expressed of a close relationship between literary history and the history of lexicography when the latter is perceived as a number of text fragments. Critics have pointed out that dictionaries consist of fragments of texts and should therefore take into consideration linguistic, literary and sociological methods. This work is an important step in elucidating the *chanson* phenomenon and the enrichment of its vocabulary and further areas of research are suggested along the lines of recent critical studies.

The author provides an extensive bibliography consisting of scientific studies on song, music and poetry, the sociology of culture and philosophy, linguistic and literary theory. A comprehensive bibliography of literary works referred to is provided as well as a section consisting of a list of dictionaries relating to musical and literary vocabulary. Ample footnotes are given, inviting further elucidation and study. This publication is not only a valuable contribution to an analysis of the *chanson*, but contains material which could be put to excellent use by future researchers.

François Jaques
Emeritus Professor of French
Rhodes University, Grahamstown, South Africa
(f.jaques@ru.ac.za)

Elektroniese WAT (Woordeboek van die Afrikaanse Taal, A tot R). CD-ROM. ISBN 978-0-981-4434-1-6. Stellenbosch: Buro van die WAT. Prys: R450.

Die *Woordeboek van die Afrikaanse Taal* is die omvangrykste projek ooit rondom die Afrikaanse woordeskat. Dit dokumenteer hierdie woordeskat in sy wydste omvang — nie slegs Standaardafrikaans nie, maar ook die ander variëteite van Afrikaans. Alle vorme, hetsy dit omgangstaal, spreektaal, streektaal of standaardtaal is, word weerspieël. Vir alle moderne en kontemporêre gebruikers is dit nou ook in elektroniese formaat beskikbaar. Dit behels 'n omvattende, elektroniese verwysingsbron met gevorderde soekfunksies wat nie bloot die data van die gedrukte dele van die WAT bevat nie, maar ook 'n kerntesourus van Afrikaans as bonus het.

Toe die *Elektroniese WAT* die eerste keer die lig gesien het op 28 Maart 2003, het die WAT sy plek ingeneem langs die leidende woordeboeke van die wêreld, soos die *Oxford English Dictionary* en die *Woordenboek der Nederlandsche Taal*. Aanprysings van tóé is vandag nog net so relevant. George Weideman het die elektroniese WAT 'n hulpmiddel genoem waarsonder 'n skrywer nie kan klaarkom nie. Nicoline van der Sijs, Nederlandse leksikograaf, het die funksionaliteit daarvan soos volg beskryf: "Het werkt als een trein. Heel erg goed." Telkens wanneer 'n nuwe, gedrukte volume van die WAT dus gepubliseer word, word die *Elektroniese WAT* dienooreenkomstig bygewerk en is dan op die internet, sowel as in CD-formaat beskikbaar en bied bykomende elektroniese moontlikhede en gebruikshulp. Die *Elektroniese WAT (A tot R)* het dus verskyn na die publikasie van Deel XIII van die WAT (die letter R) in Julie 2009. Die CD-ROM installeer maklik op enige rekenaar mits die minimum stelselvereistes verreken word — spesiale instellings vir Windows 98 en vroeër is nodig.

Kortliks: die *Elektroniese WAT* is maklik toeganklik ingedeel in 'n voorwoord, gebruikersgids, toeligting, voortekste, sentrale teks (A–R) en agtertekste. Indien die plus-teken langs elkeen van dié skakels gekies word, word verdere lae inligting blootgelê. Die *Elektroniese WAT* bied die omvangryke inligting wat in 13 WAT-dele (A tot R) vervat is — die ekwivalent van meer as 8 800 gedrukte bladsye. Verder is die *Elektroniese WAT* die omvattendste weerspieëling van die Afrikaanse woordeskat wat beskikbaar is met nagenoeg 200 000 verklaarde inskrywings. 'n Gebruikersvriendelike, vinnige en effektiewe kitsoekfasiliteit vir lemmas, etimologieë, woordsoorte en etikette, is die grootste pluspunt van die *Elektroniese WAT*. Ander soorte soektogte is ook moontlik. Verder maak kleurgekodeerde inligtingstipes, in verskillende style, naslaanwerk baie eenvoudig. Met die klik van die muis op 'n skakel kan daar vinnig van een inskrywing na 'n ander, selfs tussen kruisverwysings, rondbeweeg word — en dit sonder om die afgelope soektog se resultate te verloor. Ek het gewonder hoe die *Elektroniese WAT* die situasie hanteer waar daar kruisverwys behoort te word tussen lemmas en een van die lemmas in die reeds bewerkte gedeelte val en die ander lemma in die nog onbewerkte gedeelte S–Z, bv. *taren-*

taal = *poelpetaat*. Onder die lemma *poelpetaat* was 'n hele paar weergawes van die woord, maar ook die lemma *tarentaal*. Uiteraard was *tarentaal* nie 'n aktiewe skakel nie, aangesien die letter T nog nie bewerk is nie. As 'n toets sleutel ek toe wel *tarentaal* in en kry inderdaad verwysings na *tarentaal* in die definisie van *kuifkoptarentaal* (want die letter K is reeds bewerk).

Indien iemand nie presies weet na watter woord hy moet soek nie en net 'n paar letters aan hom bekend is, kan wildekaart-soektogte onderneem word. So kan gesoute blokkiesraaiselopstellers uitoorlê word. Blokkiesraaiselinvullers wat nog net enkele letters van 'n woord het, of nie weet hoe die woord gespel word nie, kan vraagtekens gebruik om die regte woord op te soek. Indien die leidraad "snydokter" is en iemand weet nie hoe die sinoniem gespel word nie, of net enkele letters is beskikbaar, kan hy bv. "??ir?rg" in die soekskerm invul en die resultaat dui op "chirurg".

Gevorderde soektogte, bv. vir uitdrukings en frases, is ook moontlik. Die woorde en uitdrukings in hierdie woordeboek kom uit die geskrewe én gesproke taal en sluit streektaal, geselstaal en die variëteite van Afrikaans in wat uit omvattende sitaatversamelings en korpora verkry is. Voorbeeldmateriaal bestaan uit sitate en kollokasies wat kursief gedruk word. Sitate kan opgespoor word volgens onder meer die skrywer, bron of trefwoord (of kombinasies hiervan). Ek wou onlangs kyk hoeveel verwysings na Von Wielligh daar in die WAT as geheel is en kon dit met die druk van 'n knoppie uitvind — daar is selfs verwysings na hom in die voortekste.

Illustrasiesketse (in dele A–K) is verwyder uit die sentrale teks en die volbladillustrasies is saamgegroepeer in die agtertekste van die woordeboek. Daar is leersame, dikwels volkleurillustrasies, soms met byskrifte, wat onderwerpe soos die volgende dek: Suid-Afrikaanse voëls, insekte, houtsoorte, gifplante, skoelappers, motte, skulpe, krappe en krewes, klipwerktuie, wildsbokke, edel- en halfedelgesteentes, en talle meer. Illustrasies is ingevoeg ter aanvulling van die woorddefinisies.

Kitswenke om basiese soektogte te doen is maklik, in drie stappe.

Om die betekenis van 'n woord (A tot R) te vind:

Stap 1: Sleutel [Ctrl en 1] in.

Stap 2: Tik die woord by [*Soek vir Lemma*] in.

Stap 3: Druk [OK].

Om te sien hoeveel maal en waar 'n bepaalde woord (A–Z) in die WAT voorkom:

Stap 1: Druk die snelskakel [F2].

Stap 2: Tik die woord in by die teksvenster.

Stap 3: Druk [OK]. (Tydens die proses kan daar in die resultaatuitleg die hoeveelheid kere wat die lemma voorkom, gesien word. Daar kan ook met "Volgende resultaat" en "Vorige resultaat" tussen soekresultate genavigeer word.)

Om 'n frase te soek:

Stap 1: Druk die snelskakel [F2].

Stap 2: Tik die frase tussen aanhalingstekens in by die teksvenster.

Stap 3: Druk [OK].

As bonus is 'n faksimileeweergawe van die bron, *Woordkeusegids, 'n kerntesourus van Afrikaans*, bygevoeg. Ook die *Woordkeusegids* is ingedeel in 'n voorwoord, gebruikstoeligting, afkortingslys en die alfabetiese lys (in hierdie geval van A-Z). Alle verwysings is aktiewe skakels. Die woordkeusegids voorsien die gebruiker nie net van sinonieme nie, maar ook woorde wat logieserwys aan die lemma verwant is en dus in dieselfde begrips- of saakkategorie as die lemma hoort, bv.

fiets s.nw. 1 rywiël, baiesukkel (skertsend), martelpyp (skertsend),
2 trapfiets, skopfiets, kragfiets, motorfiets, tandem, tandemfiets, eenwielfiets,
driewiel, mansfiets, damesfiets, kinderfiets, resiesfiets, baanfiets, veldfiets,
bergfiets, padfiets, sportfiets, toerfiets.

'n Toekomsprospektief: met die volgende uitgawe van die *Elektroniese WAT* sal die eerste van drie gedeeltes vir die letter S bygevoeg word. Ten spyte van 'n afgeskaalde staatsubsidie, verrig die Buro van die WAT 'n enorme taak op die terrein van die Afrikaanse woordeskat. Dit bly so kontemporêr dat die Afrikaanse leksikografie wêreldwyd as toonaangewend beskou word.

Die *Elektroniese WAT (A tot R)* op CD-ROM kos R450 en indien iemand reeds die *Elektroniese WAT (A tot Q)* besit, betaal hy R150 vir die volgende deel se bywerking. Die CD kan gekoop word by die WAT te Banghoekweg 115, Stellenbosch, of bestel word by 021 887 3113 of wat@sun.ac.za, of deur 'n bestelvorm by www.wat.co.za te kry. Gebruikers kan as alternatief inteken op die Internet-weergawe, teen 'n koste van R150 per jaar, by <http://www.woordeboek.co.za>. Geregistreerde gebruikers bly ingelig aangaande bywerkings en kry selfs afslag daarop en, indien benodig, produkondersteunings-hulp.

Corlietha Swart
Departement Afrikaans en Nederlands, Duits en Frans
Universiteit van die Vrystaat
Bloemfontein
Suid-Afrika
(swartc@ufs.ac.za)

Publikasieaankondigings / Publication Announcements

Henri Béjoint. *The Lexicography of English: From Origins to Present*. 2010, xxiv + 458 pp. ISBN 9787-0-19-829967-7 (Hardback). Oxford/NewYork: Oxford University Press. Price: £30.

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Dertiende Deel: R*. 2009, xxi + 541 pp. ISBN-13: 978-0-9584971-7-6 (plastiekband), ISBN-13: 978-0-9584971-8-3 (leerband). Stellenbosch: Buro van die WAT. Prys: R290 (plastiekband) / R590 (leerband). (Resensie in hierdie nommer.)

G.-M. de Schryver (Editor). *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. 2010, viii + 376 pp. ISBN 978-9970-101-01-6 (Hardback). Linguistics Series. Kampala: Menha Publishers. Price: €59.95. (Review in this issue.)

G.-M. de Schryver (Editor). *Oxford Bilingual School Dictionary: Zulu and English*. First Edition. 2010, lviii + 582 pp. ISBN-13: 978 0 19 576554 0 (Paperback). Cape Town: Oxford University Press Southern Africa. Price: R104.95. (Review in this issue.)

Jean Nicolas De Surmont. *Chanson: son histoire et sa famille dans les dictionnaires de langue française*. 2010, IX + 248 pp. ISBN 978-3-484-52353-1 (Hardback). Berlin/New York: Walter de Gruyter. Price: €49.95. (Review in this issue.)

Elektroniese WAT (Woordeboek van die Afrikaanse Taal, A tot R). CD-ROM. ISBN 978-0-981-4434-1-6. Stellenbosch: Buro van die WAT. Prys: R450. (Resensie in hierdie nommer.)

Pedro A. Fuertes-Olivera (Editor). *Specialised Dictionaries for Learners*. 2010, IX + 237 pp. ISBN 978-3-11-023132-8 (Hardback). Lexicographica. Series Maior 136. Berlin/New York: Walter de Gruyter. Price: €99.95

Jana Luther (Redakteur), Liezl Potgieter, Elmarie van Niekerk, Herman Beyer (Hulpredakteurs), Frikkie Lombard, Daniel Hugo (Medewerkers). *HAT Afrikaanse sakwoordeboek*. Eerste uitgawe, eerste druk. 2009, xi + 708 pp. ISBN 978-1-77025-781-8 (sagteband). Kaapstad: Pearson Education South Africa.

Jana Luther (Redakteur), Liezl Potgieter, Elmarie van Niekerk, Herman Beyer (Hulpredakteurs), Frikkie Lombard, Daniel Hugo (Medewerkers). *HAT Afrikaanse skoolwoordeboek*. Eerste uitgawe, eerste druk. 2009, xi + 708 pp.

ISBN 978-1-77025-595-1 (sagteband). Kaapstad: Pearson Education South Africa.

Sandro Nielsen and Sven Tarp (Editors). *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*. 2009, XI + 341 pp. ISBN 978 90 272 2336 4 (Hardback). Terminology and Lexicography Research and Practice 12. Amsterdam/Philadelphia: John Benjamins. Price: €99.

Anton F. Prinsloo. *Spreekwoorde en waar hulle vandaan kom*. Tweede hersiene uitgawe, eerste druk. 2009, x + 605 pp. ISBN 978-1-86890-101-2 (sagteband). Kaapstad: Pharos Woordeboeke. Prys: R325.

Bo Svensén. *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. 2009, xvi + 535 pp. ISBN 978-0-521-88180-7 (Hardback) / 978-0-521-70824-1 (Paperback). Cambridge: Cambridge University Press. Price: £70.00 (Hardback)/£29.99 (Paperback). (Review article in this issue.)

Zola Wababa, Keith Welman, Karen Press (Editors). *Isichazi-magama seziBalo Sezikolo sase-Cambridge*. 2010, iv + 203 pp. ISBN 978-0-521-13010-3 (Paperback). Cape Town: Cambridge University Press.

Herbert Andreas Welker. *Panorama geral da lexicografia pedagógica*. 2008, 519 pp. ISBN 978-85-7062-817-6 (Paperback). Brasília: Thesaurus Editora. Price: R\$ 85. (Review article in this issue.)

Herbert Ernst Wiegand, Michael Beißwenger, Rufus H. Gouws, Matthias Kammerer, Angelika Storrer, Werner Wolski (Herausgeber und Bearbeiter/Editors and Compilers) unter Mitarbeit von/with the Collaboration of Ekaterina Butina-Koller (Russisch/Russian), Rute Costa (Portugiesisch/Portuguese), M^a Theresa Fuentes Morán (Spanisch/Spanish), Laura Giacomini (Italienisch/Italian), Rufus H. Gouws (Afrikaans; Englisch/English), Franz Josef Hausmann und Maria Hegner (Französisch/French), Regina Hessky und Zita Hollós (Ungarisch/Hungarian), Pavel Petkov (Bulgarisch/Bulgarian), Giovanni Rovere (Italienisch/Italian), Stefan J. Schierholz (Portugiesisch/Portuguese), Maria Smit (Englisch/English). *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen / With English Translations of the Outer Texts and Definitions as well as Equivalents in Nine Languages. 1. Bd.: Systematische Einführung / Vol. I: Systematic Introduction. A–C. Mit 120 Grafiken von/With 120 Figures by Matthias Kammerer*. 2010, LXIX + 751 pp. ISBN 978-3-11-016472-5 (Hardback). Berlin/New York: Walter de Gruyter. [Mit/With CD-ROM.] Price: €349.

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.wat.co.za>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguïstiek, algemene taalwetenskap, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder enigeen van die volgende rubrieke geklassifiseer word:

(1) **Artikels:** Grondige oorspronklike wetenskaplike navorsing wat gedoen en die resultate wat verkry is, of bestaande navorsingsresultate en ander feite wat op 'n oorspronklike wyse oorsigtelik, interpreterend, vergelykend of krities evalueerend aangebied word.

(2) **Resensieartikels:** Navorsingsartikels wat in die vorm van 'n kritiese resensie van een of meer gepubliseerde wetenskaplike bronne aangebied word.

Bydraes in kategorieë (1) en (2) word aan streng anonieme keuring deur onafhanklike akademiese vakgenote onderwerp ten einde die internasionale navorsingsgehalte daarvan te verseker.

(3) **Resensies:** 'n Ontleding en kritiese evaluering van gepubliseerde wetenskaplike bronne en produkte, soos boeke en rekenaarprogramme.

(4) **Projekte:** Besprekings van leksikografiese projekte.

(5) **Leksikonotas:** Enige artikel wat praktykgerigte inligting, voorstelle, probleme, vrae, kommentaar en oplossings betreffende die leksikografie bevat.

(6) **Leksikovaria:** Enigeen van 'n groot verskeidenheid artikels, aankondigings en nuusvrystellings van leksikografiese verenigings wat veral vir die praktiserende leksikograaf van waarde sal wees.

(7) **Verslae:** Verslae van konferensies en werksessies.

Bydraes in kategorieë (3)-(7) moet almal aan die eise van akademiese geskrifte voldoen en word met die oog hierop deur die redaksie gekeur.

2. Wetenskaplike standaard en keuringsprosedure

Lexikos is deur die Departement van Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde d.w.s. inkomstegenererende navorsingstydskrif goedgekeur.

Artikels sal op grond van die volgende aspekte beoordeel word: taal en styl; saaklikheid en verstaanbaarheid; probleemstelling, beredenering en gevolgtrekking; verwysing na die belangrikste en jongste literatuur; wesenlike bydrae tot die spesifieke vakgebied.

3. Taal van bydraes

Afrikaans, Duits, Engels, Frans of Nederlands.

4. Kopiereg

Nóg die Buro van die WAT nóg die African Association for Lexicography (AFRILEX) aanvaar enige aanspreeklikheid vir eise wat uit meewerkende skrywers se gebruik van materiaal uit ander bronne mag spruit.

Outeursreg op alle materiaal wat in *Lexikos* gepu-

bliseer is, berus by die Direksie van die Woordeboek van die Afrikaanse Taal. Dit staan skrywers egter vry om hulle materiaal elders te gebruik mits *Lexikos* (AFRILEX-reeks) erken word as die oorspronklike publikasiebron.

5. Oorspronklikheid

Slegs oorspronklike werk sal vir opname oorweeg word. Skrywers dra die volle verantwoordelikheid vir die oorspronklikheid en feitelike inhoud van hulle publikasies.

6. Gratis oordrukke en eksemplare

Skrywers ontvang vyf gratis oordrukke van elke artikel of resensieartikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evalueerende resensies en van bydraes tot die rubrieke Leksikonotas, Leksikovaria, Projekte en Verslae ontvang vyf gratis oordrukke van hulle bydraes. In laasgenoemde vier kategorieë kan die redaksie egter, afhangend van die aard en omvang van die bydraes, besluit om ook 'n eksemplaar van die betrokke uitgawe aan 'n skrywer toe te ken.

7. Uitnodiging en redaksionele adres

Alle belangstellende skrywers is welkom om bydraes vir opname in *Lexikos* te lewer en aan die volgende adres te stuur:

Die Redakteur: LEXIKOS

Buro van die WAT

Posbus 245

7599 STELLENBOSCH

Republiek van Suid-Afrika

B. VOORBEREIDING VAN MANUSKRIP

Die manuskrip van artikels moet aan die volgende redaksionele vereistes voldoen:

1. Lengte en formaat van artikels

Bydraes moet verkieslik nie 20 getikte A4-bladsye met teks in dubbelspasiëring en ruim kantlyne (ongeveer 2,5 cm) oorskry nie. Manuskrip moet verkieslik in elektroniese formaat as ASCII-teks, as volledig geformateerde Microsoft Word (DOS of Windows) lêers of as WordPerfect (DOS of Windows) lêers op rekenaar-skyf (360 KB tot 1.44 MB) voorgelê word. 'n Rekenaardrukstuk van die artikel moet die skyf vergesel. Elke artikel moet voorsien wees van 'n Engelse opsomming van tussen 150 en 250 woorde, sowel as tussen 10 en 30 Engelse sleutelwoorde.

2. Grafika

Een stel duidelike oorspronklike illustrasies, tabelle, grafieke, diagramme, of kwaliteitsafdrukke daarvan, moet voorgelê word. Die plasing van grafika binne die teks moet duidelik aangedui word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za)
or refer to our website: <http://www.wat.co.za>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may deal with pure lexicography or with the implications that related fields such as linguistics, general linguistics, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research that has been done and the results that have been obtained, or reflecting existing research results and other facts in an original, synoptic, interpretative, comparative or critically evaluative manner.

(2) **Review articles:** Research articles presented in the form of a critical review of one or more published scientific sources.

Contributions in categories (1) and (2) are subjected to strict anonymous evaluation by independent academic peers in order to ensure the international research quality thereof.

(3) **Reviews:** An analysis and critical evaluation of published scientific sources and products, such as books and computer software.

(4) **Projects:** Discussions of lexicographical projects.

(5) **Lexiconotes:** Any article containing practice-oriented information, suggestions, problems, questions, commentary and solutions regarding lexicography.

(6) **Lexicovaria:** Any of a large variety of articles, announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Reports:** Reports on conferences and workshops.

Contributions in categories (3)-(7) must all meet the requirements of academic writing and are evaluated by the editors with this in mind.

2. Academic standard and evaluation procedure

The Department of Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal.

Articles will be evaluated on the following aspects: language and style; conciseness and comprehensibility; problem formulation, reasoning and conclusion; references to the most important and most recent literature; substantial contribution to the specific discipline.

3. Language of contributions

Afrikaans, Dutch, English, French or German.

4. Copyright

Neither the Bureau of the WAT nor the African Association for Lexicography (AFRILEX) accepts any responsibility for claims which may arise from contributing authors' use of material from other sources.

Copyright of all material published in *Lexikos* will be vested in the Board of Directors of the Woordeboek van die Afrikaanse Taal. Authors are free however to use their material elsewhere provided that *Lexikos* (AFRILEX Series) is acknowledged as the original publication source.

5. Originality

Only original contributions will be considered for publication. Authors bear full responsibility for the originality and factual content of their contributions.

6. Free offprints and copies

Authors will receive five free offprints of each of their articles or review articles published, as well as one complimentary copy of the issue containing such article(s). Authors of purely evaluative reviews and of contributions to the categories Lexiconotes, Lexicovaria, Projects, and Reports receive five free offprints of their contributions. In the case of the latter four categories, the editors may, however, depending on the nature and scope of the contributions, decide to grant the author a copy of the issue concerned.

7. Invitation and editorial address

All interested authors are invited to submit contributions for publication in *Lexikos* to:

The Editor: LEXIKOS
Bureau of the WAT
P.O. Box 245
7599 STELLENBOSCH
Republic of South Africa

B. PREPARATION OF MANUSCRIPTS

Manuscripts of articles must meet the following editorial requirements:

1. Length and format

Contributions should not exceed more than 20 type-written A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte uns Kontakt auf (lexikos@sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: <http://www.wat.co.za>)

A. REDAKTIONELLE ZIELSETZUNGEN

1. Art und Inhalt der Artikel

Es können Artikel aufgenommen werden, die sich mit Themen der Lexikographie befassen oder mit Zusammenhängen, die zwischen der Lexikographie und benachbarten Fachgebieten wie z.B. Linguistik, allgemeiner Sprachwissenschaft, Lexikologie, Computerwissenschaft und Management bestehen.

Die Beiträge sollten einer der folgenden Kategorien entsprechen:

(1) **Artikel**, die grundlegend über neue Forschungsansätze und deren Ergebnisse berichten, oder die bestehende Forschungsergebnisse und andere Informationen selbständig, interpretativ, vergleichend oder kritisch bewertend wiedergeben.

(2) **Rezensionsartikel**, die in der Form eines Forschungsartikels eine oder mehrere veröffentlichten wissenschaftlichen Quellen kritisch rezensieren.

Beiträge in Kategorien (1) und (2) werden streng anonym von unabhängigen wissenschaftlichen Experten begutachtet, um ein internationales fachliches Niveau in *Lexikos* zu gewährleisten.

(3) **Rezensionen**, die veröffentlichte wissenschaftliche Quellen und Produkte, wie z.B. Bücher und Software, analysieren und kritisch bewerten.

(4) **Lexikographische Projekte**, die vorgestellt werden.

(5) **Notizen zum Lexikon**, die praxisbezogene Informationen, Vorschläge, Probleme, Fragen, Kommentare und Lösungen hinsichtlich der Lexikographie enthalten.

(6) **Lexikovaria**, die unterschiedliche Beiträge, Ankündigungen und Pressemitteilungen lexikographischer Vereinigungen, die dem praktischen Lexikographen wichtig sein können, einschließen.

(7) **Berichte** über Konferenzen und Workshops.

Beiträge in Kategorien (3)-(7) müssen im akademischen Stil abgefaßt werden. Sie werden von der Redaktion unter diesem Gesichtspunkt beurteilt.

2. Wissenschaftliche Standards und das Beurteilungsverfahren

Das Erziehungsministerium der südafrikanischen Regierung hat *Lexikos* als eine subventionierte, d.h. einkommenerzeugende Forschungszeitschrift anerkannt.

Artikel werden auf Grund der folgenden Gesichtspunkte bewertet: Sprache und Stil; Sachlichkeit und Verständlichkeit; Problembeschreibung, Argumentation und Schlußfolgerung; Hinweise auf die neueste und wichtigste Literatur; wesentlicher Beitrag zum besonderen Fachgebiet.

3. Sprache der Beiträge

Afrikaans, Deutsch, Englisch, Französisch oder Niederländisch.

4. Das Urheberrecht

Weder das Büro des WAT noch die African Association for Lexicography (AFRILEX) übernehmen Verantwortung für Ansprüche, die daraus entstehen könnten, daß Autoren Material aus anderen Quellen benutzt haben.

Das Urheberrecht aller in *Lexikos* publizierten Artikel wird dem Direktorium unseres Büros übertragen. Es steht Autoren jedoch frei, ihren Beitrag anderweitig zu verwenden, vorausgesetzt, *Lexikos* (AFRILEX-Serie) wird als Originalquelle genannt.

5. Originalität

Nur Originalbeiträge werden begutachtet. Autoren tragen die volle Verantwortung für die Originalität und den sachlichen Inhalt ihrer Beiträge.

6. Sonderdrucke und Freixemplare

Autoren erhalten fünf Sonderdrucke ihrer veröffentlichten Artikel oder Rezensionsartikel gratis sowie ein Freixemplar der betreffenden Ausgabe. Rezensenten und Autoren von Beiträgen zu den Kategorien Notizen zum Lexikon, Lexikovaria, Projekte und Berichte erhalten fünf Sonderdrucke ihrer Beiträge gratis. Die Redaktion kann sich jedoch, abhängig von der Art und dem Umfang der Beiträge der letztgenannten vier Kategorien, vorbehalten, dem Autor ein Freixemplar der Ausgabe zu überlassen.

7. Einladung und redaktionelle Adresse

Alle Autoren, die interessiert sind, Beiträge für *Lexikos* zu liefern, sind herzlich willkommen. Sie werden gebeten, ihre Artikel an die folgende Adresse zu schicken:

Der Redakteur: LEXIKOS
Buro van die WAT
Postfach 245
7599 STELLENBOSCH
Republik Südafrika

B. VORBEREITUNG DES MANUSKRIPTS

Ein Artikelmanuskript muß den folgenden redaktionellen Anforderungen entsprechen:

1. Umfang und Format

Beiträge sollen nicht länger als 20 getippte A4-Seiten in zweizeiligem Abstand und mit Randabständen von ca. 2,5 cm sein. Das Manuskript sollte möglichst als elektronischer Text auf einer (360 KB bis 1.44 MB) Diskette vorgelegt werden, entweder im ASCII-Format, oder in formatiertem Microsoft Word (DOS oder Windows) bzw. WordPerfect (DOS oder Windows). Ein Ausdruck des vollständig formatierten Artikels soll mit der Diskette eingereicht werden. Jedem Artikel ist eine Zusammenfassung im Umfang von 150-250 Wörtern beizufügen. Ferner sollen etwa 10-30 inhaltskennzeichnende Stichwörter zu jedem Artikel angegeben werden.

2. Abbildungen

Ein reproduktionsfähiger Satz der originalen Abbildungen, Illustrationen, Tabellen, Graphiken und Diagramme oder Qualitätsabdrucke muß vorgelegt werden. Der Text selber sollte klare Hinweise auf die Position der Abbildungen enthalten.

3. Bibliographische Einzelheiten und Hinweise im Text

Zu Einzelheiten des bibliographischen Systems sind neuere Ausgaben von *Lexikos* einzusehen.

INSTRUCTIONS AUX AUTEURS

(Pour une version plus détaillée de ces instructions, contacter le Bureau du WAT (lexikos@sun.ac.za)
ou consulter notre website: <http://www.wat.co.za>)

A. POLITIQUE ÉDITORIALE

1. Caractéristiques et contenu des articles

Les articles seront consacrés à la lexicographie pure, ou aux rapports entre la lexicographie et les disciplines voisines telles que la linguistique, la linguistique générale, l'informatique et le management.

Les contributions pourront appartenir à l'une des catégories suivantes:

- (1) **Articles:** Recherches scientifiques originales, avec leurs résultats; ou présentations originales, synoptiques, interprétatives, comparatives, évaluatives et critiques des résultats de recherches en cours;
- (2) **Articles bilans:** Articles de recherche présentés sous forme de bilan critique de travaux scientifiques déjà publiés.

Les contributions appartenant aux catégories (1) et (2) seront soumises de manière anonyme à des experts spécialistes indépendants afin d'en assurer la qualité scientifique au niveau international.

- (3) **Recensions:** Analyses et évaluations critiques de travaux de recherche et de productions scientifiques, telles que livres ou logiciels;
- (4) **Projets:** Présentations de projets lexicographiques;
- (5) **'Lexiconotes':** Textes contenant des informations pratiques, ou des suggestions, des problèmes, des questions, des commentaires et des solutions concernant des activités lexicographiques;
- (6) **'Lexicovaria':** Articles, annonces, communiqués de presse émanant de centres de lexicographie et qui revêtent un intérêt particulier pour les lexicographes;
- (7) **Rapports:** Rapports sur des colloques et ateliers.

Les contributions dans les catégories (3) à (7) devront répondre aux exigences de qualité des publications scientifiques et seront évaluées dans cette optique.

2. Critères et procédures d'évaluation

La revue *Lexikos* est reconnue et subventionnée par le Ministère de l'Éducation du gouvernement Sud-Africain comme revue devant générer des revenus.

Les articles seront évalués selon les critères suivants: langue et style, concision et clarté, formulation de la problématique, raisonnement et conclusion, référence aux travaux les plus importants et les plus récents, contribution substantielle à la discipline.

3. Langue des contributions

Afrikaans, allemand, anglais, français ou néerlandais.

4. Copyright

Le Bureau du WAT ou l'*African Association for Lexicography* (AFRILEX) décline toute responsabilité en cas de réclamations motivées par l'utilisation d'autres sources par les auteurs.

Les droits d'auteurs des documents publiés dans *Lexikos* appartiennent au Conseil d'administration du *Woordboek van die Afrikaanse Taal* (WAT). Cependant,

les auteurs sont libres d'utiliser leurs textes dans d'autres publications, à condition d'indiquer *Lexikos* (collection AFRILEX) comme source de la publication originale.

5. Originalité

Seules les contributions originales seront acceptées pour la publication. Les auteurs conservent l'entière responsabilité de l'originalité et du contenu factuel de leur texte.

6. Tirés-à-part et exemplaires gratuits

Les auteurs recevront gratuitement cinq (5) tirés-à-part de chaque article ou article de bilan, ainsi qu'un exemplaire gratuit de la publication contenant cet article.

Les auteurs des recensions et des publications dans les catégories 'Lexiconotes', 'Lexicovaria', Projets et Rapports recevront cinq (5) tirés-à-part de leur texte. Dans ces quatre dernières catégories, le responsable de la revue pourra néanmoins décider, en fonction de la nature et des dimensions des publications, d'accorder un exemplaire gratuit de la revue à leurs auteurs.

7. Adresse de la revue

Les auteurs intéressés sont invités à soumettre leurs propositions à:

L'éditeur: LEXIKOS
Bureau du WAT
Boîte postale 245
7599 STELLENBOSCH
République d'Afrique du Sud

B. PRÉSENTATION DES MANUSCRITS

Les manuscrits se conformeront aux exigences suivantes:

1. Longueur et format

Les contributions ne devraient pas excéder 20 pages dactylographiées, de format A4, avec double espacement et marges suffisantes (environ 2,5 cm), si possible en format électronique (disquette 360KB à 1,44MB), sous forme de fichier ASCII, complètement formaté sous Microsoft Word ou sous WordPerfect (DOS ou Windows). La disquette sera accompagnée d'un tirage papier. Chaque article sera pourvu d'un résumé en anglais de 150 à 200 mots, et de 10 à 30 mots-clés.

2. Tableaux et graphiques

Les dessins, tableaux, graphiques et diagrammes seront envoyés, soit sous leur forme originale soit sous forme d'une copie de bonne qualité. Leur place dans le texte devra être clairement indiquée.

3. Bibliographie et références dans le texte

Voir les exemplaires récents de *Lexikos*.