

Lexikos 18

Lexikos 18

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 18:2008

AFRILEX SERIES 18:2008



BURO VAN DIE WAT

STELLENBOSCH

Die publikasie van hierdie boek is moontlik gemaak deur 'n ruim subsidie 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 subsidy 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 © 2008 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2008

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-13: 978-0-9584971-6-9
ISBN-10: 0-9584971-6-8
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

'n Elektroniese weergawe van *Lexikos* is beskikbaar by Sabinet Online
An electronic version of *Lexikos* is available from Sabinet Online

Indekse Indexes

Arts & Humanities Citation Index, Current Contents/Arts & Humanities, Elsevier: Scopus, R.R.K. Hartmann Bibliography of Lexicography, South African Bibliographic and Information Network, SIL Bibliography Serials Index

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)
Prof. J. Taeldeman (België/Belgium)
Dr. J. Tent (Australië/Australia)
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)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	x
Foreword	xi
<i>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xii
A Few Words from AFRILEX	xiii
<i>Rufus H. Gouws</i>	
Redaksionele doelstellings	xiv
Editorial Objectives	xv
Redaktionelle Ziele	xvi
Politique éditoriale	xvii

Artikels / Articles

The Ideology of the Perfect Dictionary: How Efficient Can a Dictionary Be?	1
<i>Michaël Abecassis</i>	
The Effectiveness of Dictionary Examples in Decoding: The Case of Kuwaiti Learners of English	15
<i>Hashan Al-Ajmi</i>	
The Dilemma of Grammatical Data in Travel Dictionaries	27
<i>Birger Andersen and Patrick Leroyer</i>	
Voorstelle vir die produksie van 'n vertaalwoordeboek vir professionele vertalers	46
<i>M.F. Crafford en Ilse Feinauer</i>	
A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary	63
<i>Gilles-Maurice de Schryver</i>	
The Lexicographic Treatment of Quantitative Pronouns in Zulu	92
<i>Gilles-Maurice de Schryver</i>	

Bilingual Dictionaries, the Lexicographer and the Translator <i>Rachéle Gauton</i>	106
Standardization or Stigmatization? Challenges Confronting Lexicography and Terminography in Sesotho sa Leboa <i>V.M. Mojela</i>	119
Lexicography and Linguistic Creativity <i>Rosamund Moon</i>	131
Traduire les énoncés religieux indigénisés: cas de la traduction du cilubà en français et en anglais <i>Germain Mulamba Nshindi</i>	154
The Effect of Lexicographical Information Costs on Dictionary Making and Use <i>Sandro Nielsen</i>	170
'n Moontlike model vir die bewerking van idiome in tweetalige woordeboeke <i>Liezl Potgieter</i>	190
Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria <i>Piet Swanepoel</i>	207
A Lexicographic Approach to Language Policy and Recommen- dations for Future Dictionaries <i>Sven Tarp and Rufus H. Gouws</i>	232
Wörterbuchregister. Grundlagen einer Theorie der Register in modernen Printwörterbüchern <i>Herbert Ernst Wiegand</i>	256
Lexikoprogrammatuur / Lexicosoftware	
Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes <i>Guy De Pauw and Gilles-Maurice de Schryver</i>	303

Projekte / Projects

- Two Recent Major Afrikaans–English/English–Afrikaans
Dictionaries from Pharos 319
Fred Pheiffer
- Wine and Words: A Trilingual Wine Dictionary for South Africa 337
Michele F. van der Merwe

Leksiko-opname / Lexicosurvey

- A Survey of the Dictionary Use of Gabonese Students 349
at Two South African Universities
Guy-Modeste Ekwa Ebanéga and Fatima Tomba Moussavou

Leksikonota / Lexiconote

- L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography 366
Munzhedzi James Mafela

Leksikobibliografie / Lexicobibliography

- Bibliographical Materials for Afrikaans Etymological 374
Lexicography
Jeremy Bergerson

Leksikonuus / Lexiconews

- Un vent nouveau pour les éditions Robert 410
Jean-Nicolas De Surmont

Leksiko-eerbewys / Lexicohonour

- B.T. Sue Atkins Receives an Honorary Doctorate from the 412
University of Pretoria

Leksikohuldeblyk / Lexicotribute

- D.C. Hauptfleisch (14 Maart 1926–14 Junie 2008) 417
D.C. Hauptfleisch (14 March 1926–14 June 2008) 420

Resensieartikel / Review Article

- An Analysis of *The Oxford Guide to Practical Lexicography* (Atkins and Rundell 2008) 423
Gilles-Maurice de Schryver

Resensies / Reviews

- Anna Dziemianko: *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English* 446
Michaël Abecassis
- Henrik Gottlieb and Jens Erik Mogensen (Editors): *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004* 449
Maria Smit
- Paul Newman: *A Hausa–English Dictionary* 456
Malami Buba

Publikasieaankondigings / Publication Announcements

- Voorskrifte aan Skrywers 465
Instructions to Authors 466
Richtlinien für Autoren 467
Instructions aux Auteurs 468

Voorwoord

Die eerste uitgawe van *Lexikos* is aan mnr. D.C. Hauptfleisch by sy aftrede as hoofredakteur van die *Woordeboek van die Afrikaanse Taal* (WAT) aangebied. Aangesien dit bedoel was as 'n eenmalige huldigingsbundel, het dit derhalwe geen uitgawenommer nie. Dit is eers daarna dat besluit is om daarmee voort te gaan as 'n jaarlikse tydskrif. Nou, 18 jaar later, word hierdie nommer van *Lexikos* aan die gedagtenis van mnr. Hauptfleisch opgedra wat op 14 Junie vanjaar oorlede is. 'n Huldeblyk aan hom verskyn op bladsye 417-419.

Behalwe die gewone afdelings waarin die inhoud van *Lexikos* verdeel word, bevat hierdie nommer ook 'n aantal nuwe afdelings soos "Leksiko-opname" en "Leksikobibliografie" om die verskillende navorsingsgebiede te dek wat in *Lexikos* behandel word.

Daar is twee veranderinge aan die Redaksiekomitee. Prof. E.F. Kotzé was 'n lid vir dertien jaar en prof. L.J. Louwrens vir vyf jaar. Al twee het 'n belangrike rol gespeel in die uitbouing van *Lexikos* as 'n leksikografiese tydskrif. Dit is met groot leedwese dat *Lexikos* van prof. Louwrens se siekte en, vroeër vanjaar, sy dood verneem het.

Lexikos wil twee nuwe Redaksiekomiteeledes, dr. V.M. Mojela en dr. H.S. Ndinga-Koumba-Binza, verwelkom, en die wens uitspreek dat hulle 'n aangename en verrykende verbintenis met *Lexikos* sal hê.

Die bydraers moet bedank word vir die voorlegging van gehalteartikels, en die keurders vir die kontrolering van die gehalte van hierdie bydraes.

'n Groot dankie word gerig aan Me. Riette Ruthven vir haar bedrewe en bekwame set van hierdie nommer. 'n Dankwoord gaan ook aan Me. Tanja Harteveld wat die rekenaars en rekenaarsstelsel in 'n werkende toestand gehou het.

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

Foreword

The first issue of *Lexikos* was presented to Mr D.C. Hauptfleisch on his retirement as editor-in-chief of the *Woordeboek van die Afrikaanse Taal* (WAT). As it was meant as a unique festschrift, it therefore has no issue number. It was only afterwards that it was decided to continue it as an annual journal. Now, 18 years later, this issue of *Lexikos* is dedicated to the memory of Mr Hauptfleisch who died on 14 June this year. A tribute to him appears on pages 420-422.

In addition to the usual sections into which the contents of *Lexikos* are divided, this issue also contains some new sections such as 'Lexicosurvey' and 'Lexicobibliography' to cover the diverse lexicographical research areas treated in *Lexikos*.

There have been two changes to the Editorial Board. Prof. E.F. Kotzé was a member for thirteen years and Prof. L.J. Louwrens for five years. Both played an important part in the development of *Lexikos* as a lexicographical journal. It was with much regret that *Lexikos* learnt of Prof. Louwrens's illness and, earlier this year, his decease.

Lexikos wants to welcome two new Editorial Board members, Dr V.M. Mojela and Dr H.S. Ndinga-Koumba-Binza, and express the wish that they will have a pleasant and enriching attachment to *Lexikos*.

The contributors must be thanked for submitting quality articles, and the referees for verifying the quality of these contributions.

A great thank you is extended to Ms Riette Ruthven for her accomplished and efficient typesetting of this issue. A word of thanks also goes to Ms Tanja Harteveld who kept the computers and the computing system in running order.

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

'n Woord van AFRILEX

Volgens 'n ou Kaapse gesegde is 'n vakvereniging sonder 'n eie tydskrif soos 'n spook sonder 'n laken. AFRILEX is dankbaar dat *Lexikos* aan ons vereniging volwaardige en omvattende spookstatus gee. Sedert sy stigting in 1995 is AFRILEX gekenmerk as 'n vereniging wat verskeidenheid vertoon. Dit is merkbaar aan sy lede wat 'n mengsel van teoretiese leksikograwe, praktiese leksikograwe en ander leksikografies geïnteresseerdes is, afkomstig uit Suid-Afrika, elders uit Afrika en uit verskeie lande buite Afrika. Die verskeidenheid is ook sigbaar wat betref die inhoud van die jaarlikse kongresprogramme waar onderwerpe van uiteenlopende aard bespreek word. Die aard van die leksikografiese bedrywighede van ons lede, die projekte uit alle sferes van die leksikografiese praktyk en die verskillende, en dikwels opponerende, teoretiese denkrigtings vertoon ook hierdie verskeidenheid. Dit is vir AFRILEX belangrik dat ons tydskrif 'n soortgelyke verskeidenheid moet weerspieël. Vanjaar se *Lexikos* doen presies dit en doen dit met sukses. Die temas wat in die artikels behandel word, die herkoms van die medewerkers maar ook die verskeidenheid onderafdelings in hierdie nommer van *Lexikos* ondersteun die beginsel van verskeidenheid. Hierdie verskillende onderafdelings maak dit ook moontlik dat die bestek van die leksikografiese verslaggewing aansienlik verbreed word. AFRILEX waardeer die mate waarin so baie leksikografies geïnteresseerdes bydra tot ons tydskrif maar waardeer eweneens die sorg waarmee die redakteur, dr. J.C.M.D. du Plessis, en me. Riette Ruthven daarin slaag om daardie bydraes te kies wat die publikasie van 'n tydskrif van internasionale gehalte verseker.

Die eerste nommer van *Lexikos* wat in 1991 verskyn het, is opgedra aan mnr. D.C. Hauptfleisch wat aan die einde van daardie jaar as Hoofredakteur van die *Woordeboek van die Afrikaanse Taal (WAT)* afgetree het. In die huidige nommer van *Lexikos* word daar weer eens 'n bydrae aan mnr. Hauptfleisch gewy — 'n huldeblyk hierdie keer, aangesien hy vroeër vanjaar oorlede is. AFRILEX huldig hom ook en erken graag sy bydrae tot die Afrikaanse leksikografie maar ook sy gretigheid om te help om 'n meer algemene leksikografiese gesprek in Suid-Afrika te bewerkstellig.

R.H. Gouws
President

A Few Words from AFRILEX

According to an old Cape saying a scientific association without its own journal is like a ghost without a sheet. AFRILEX appreciates the fact that *Lexikos* awards our association full and comprehensive ghost status. Since its establishment in 1995, AFRILEX has been characterised as an association displaying diversity. This is noticeable in its membership, consisting of a mixture of theoretical lexicographers, practical lexicographers and other lexicographically interested people, coming from South Africa, other countries in Africa and several countries outside Africa. The diversity can also be seen in the contents of the annual conference programmes where topics of a diverse nature are discussed. The nature of the lexicographic activities of our members, the projects from all spheres of the lexicographic practice and the different, often opposing, theoretical schools of thought also display this diversity. For AFRILEX, it is important that our journal should reflect a similar diversity. This year's *Lexikos* does exactly this and does it successfully. The topics treated in the different articles, the origin of the participating authors but also the variety of subsections in this issue of *Lexikos* support the principle of diversity. These different subsections also make it possible that the scope of the lexicographic coverage is considerably enlarged. AFRILEX appreciates the way in which so many lexicographically interested people contribute to our journal but we also appreciate the care with which the editor, Dr J.C.M.D. du Plessis, and Ms Riette Ruthven manage to select those contributions that ensure the publication of a journal of international quality.

The first issue of *Lexikos*, published in 1991, was dedicated to Mr D.C. Hauptfleisch who retired at the end of that year as Editor-in-chief of the *Woordeboek van die Afrikaanse Taal* (WAT). In the current issue of *Lexikos*, a contribution is yet again devoted to Mr Hauptfleisch — a tribute this time, because he passed away earlier this year. AFRILEX also honours him and gladly acknowledges his contribution to Afrikaans lexicography but also his eagerness to assist in establishing a more general lexicographic discussion in South Africa.

R.H. Gouws
President

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 Ideology of the Perfect Dictionary: How Efficient Can a Dictionary Be?

Michaël Abecassis, *Department of Modern Languages, University of Oxford, Oxford, United Kingdom (michael.abecassis@mod-langs.oxford.ac.uk)*

Abstract: Dictionaries have become essential tools of the modern world. Not only have dictionary sales dramatically increased, but the variety of dictionaries and the competition between editors are also very much on the rise. Monolingual dictionaries attract native speakers for several reasons. Some wish to capture the subtleties of their own language, others to speak the 'standard' language, and an 'ideologically' politically correct variety devoid of colloquialisms, hence the crucial role played by style labels. Furthermore, a large number of word enthusiasts enjoy linguistic curiosities, archaisms and other vestiges from the past conserved in dictionaries. Is the concept of a perfect dictionary a reality or an ideal? There is no perfect student. Language learners, for whom dictionaries are of great importance, seek user-friendly material which will improve both their fluency in and understanding of the target language, and embed acquired lexis in their long-term memory. Lexicographers, in their search for perfection and in compliance with users' wishes, are constantly innovating, and every dictionary hopes to become a landmark in lexicography and in second language acquisition. This article aims to look at the way dictionaries have evolved and assess the latest generation of computer-based dictionaries, as well as consider possible developments which will contribute to the compilation of future dictionaries.

Keywords: LEXICOGRAPHY, LEXICAL ACQUISITION, VOCABULARY, STYLE LABELS, CORPUS/CORPORA, DICTIONARIES, IDEOLOGY, STANDARD, LANGUAGE LEARNING, FRENCH MONOLINGUAL DICTIONARIES, ELECTRONIC DICTIONARIES, CD-ROMS

Opsomming: Die ideologie van die volmaakte woordeboek: Hoe doeltreffend kan 'n woordeboek wees? Woordeboeke het noodsaaklike werktuie van die moderne wêreld geword. Nie alleen het woordeboekverkope dramaties vermeerder nie, maar die verskeidenheid woordeboeke en die wedywering tussen redakteurs is ook aansienlik aan die toeneem. Eentalige woordeboeke trek moedertaalsprekers om verskeie redes aan. Sommige wil die subtiliteite van hul eie taal bemeester, ander wil die "standaard"-taal en 'n "ideologies" polities korrekte variëteit sonder gemeensame uitdrukkings praat, vandaar die uiters belangrike rol wat styletikette speel. Daarbenewens hou 'n groot aantal woordentoesiaste van taaleiesoortighede, argaïsmes en ander oorblyfsels uit die verlede wat in woordeboeke bewaar word. Is die konsep van 'n volmaakte woordeboek 'n realiteit of 'n ideaal? Daar is geen volmaakte student nie. Taalaanleerders vir wie woordeboeke van groot belang is, soek gebruikersvriendelike gegewens wat sowel hul vlotheid in as hul verstaan van die doeltaal sal verbeter, en aangeleerde lexis in hul langtermyngeheue vaslê. Leksikograwe, in hul soeke na volmaaktheid en ooreenkomstig gebruikers se wense, is voortdurend besig om te vernuwe en elke woordeboek hoop om 'n baken in die leksikografie en tweedetaalverwerwing te word. Hierdie artikel wil kyk na die manier waarop woordeboeke ontwikkel het en die jongste generasie rekenaargebaseerde woordeboeke beoordeel sowel as moontlike ontwikkelinge beskou wat sal bydra tot die samestelling van toekomstige woordeboeke.

Sleutelwoorde: LEKSIKOGRAFIE, LEKSIKALE VERWERWING, WOORDESKAT, STYL-ETIKETTE, KORPUS/KORPUSSE, WOORDEBOEKE, IDEOLOGIE, STANDAARD, TAALAAN- LERING, FRANSE EENTALIGE WOORDEBOEKE, ELEKTRONIESE WOORDEBOEKE, CD- ROMS

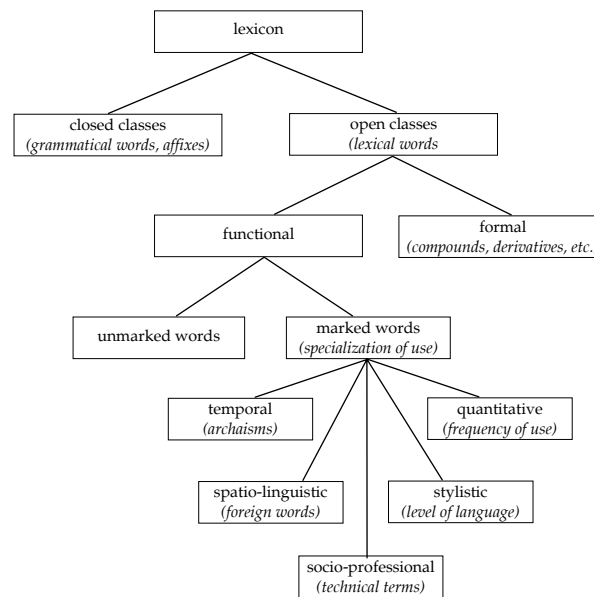
Dictionaries are commonly regarded as authoritative reference books which establish the standardness of the language, the appropriate spelling, the definition, the register or the grammatical status of a particular lexical item. They have become an essential part of people's lives with thousands being sold every year throughout the world. The concept of a standard language is in itself a subjective evaluation: it represents 'an idea in the mind rather than a reality — a set of abstract norms to which actual usage may conform to a greater or lesser extent' (Milroy and Milroy 1992: 23). 'Standard' therefore refers to the long-established codified variety, promoted by education, grammar books, text-books and dictionaries, that establish what is 'correct' and proscribe what should not be said or written. Such a system of attitudes to linguistic variation has been labelled the 'ideology of the standard' (Milroy and Milroy 1992: 23). The attitude to lexicography has changed since the 1970s with dictionaries being increasingly descriptive rather than prescriptive. However, today's general-purpose dictionaries and EFL learner's dictionaries, descriptive though they may be, still strongly emphasize the standard language (more non-standard words are included than in the past but categorized and style-labelled with a wide range of markers to account for socio-situational variation) and there may be speculated whether the image they reflect of language is accurate or purely ideological. Leech and Nesi (1999: 295) in their study of modern EFL dictionaries acknowledge the growing interest and advances in lexicography, but are forced to recognize that these dictionaries 'fall well short of perfection'. This article will concentrate mostly on EFL dictionaries, with a particular focus on French and English, and will gauge whether it is possible for lexicographers relying more and more on large-scale language corpora to give a true picture of the vitality of the language.

1. The social and stylistic continuum: the example of French monolingual dictionaries

The attitude of the speaker towards language is very complex and relies on 'fictive, prescriptive and evaluative norms' (Houdebine 1982: 50), based on extralinguistic factors which could be political, social, psychological or ideological. Although Quirk et al. (1985: 26) suggest a social and stylistic stratification of English into 'very formal–FORMAL–neutral–INFORMAL–very informal', the OALDCE lists fifteen different stylistic values that account for temporality, connotations and denotations attached to each entry. *Webster's New Collegiate Dictionary* (1976) classifies words as 'slang', 'nonstandard' and 'substandard', each style label defined by its relation to a standard or norm, whereas COBUILD

favours the terminology 'formal', 'informal', and 'offensive' words. As Stein (2002: 14) notices 'it is admittedly very difficult to make objective assessments on the social status of words, but it seems [...] that we need much more research in this area'. There is a certain degree of inconsistency in the use of style labels in dictionaries. Imbs (1969: 51) sees the role of style labels as specifying 'des niveaux ou registres de langue, c'est-à-dire des connotations de nature sociale, culturelle ou affective' ('language levels or registers, that is to say social, cultural or affective connotations'). The divide between the prestige norm (codified usage) and non-standard items (colloquial usage) in the lexicon can be expressed in terms of high and low varieties (Lodge 1989: 427-428). All languages have a stock of high-value words reserved for formal contexts and low-value words for informal situations. It is the knowledge of these different styles that constitutes learners' fluency.

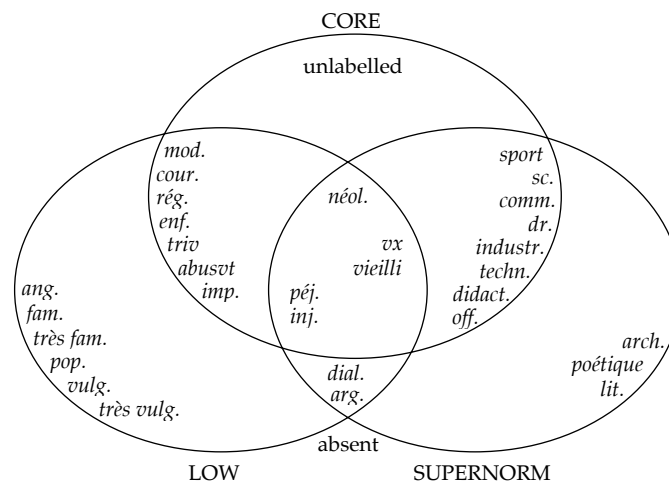
Wooldridge (1977), in his study of a nomenclature of early dictionaries, classifies the lexicon lemmatized in dictionaries in the way shown in Graph 1.



Graph 1: Nomenclature of word classes in a lexicon (adapted from Wooldridge 1977: 83 by Shaw 1997)

Low-status items, which had since the 17th century been regarded more as curiosities in France (Oudin 1640, Michel 1807), gave birth to numerous glossaries or dictionaries (Rigaud 1878, Caradec 1977, Merle 1986). They are now widely included in general dictionaries, being part of the linguistic repertoire of native speakers and often popularized by the media and a young generation willing to be subversive and politically incorrect.

Graph 2 represents the conception of the stylistic continuum in modern French dictionaries, to which have been added the designations *unlabelled* and *absent*. *Unlabelled* items constitute the stylistically unmarked core of the lexicon. *Absent* words refer on the one hand to those words at the bottom of the stylistic continuum that lexicographers do not select for inclusion or on the other hand to those words belonging to the supernorm too specialized for a general dictionary. Three subgroups are distinguished by means of circles: the low variety, the core and the 'supernorm'. The core corresponds to the 'common language' (Meillet 1921). This graphic representation of style labels therefore intends to contrast the core and the low variety (usages that infringe the prescriptive norm) with the 'supernorm' (Müller 1985: 226), for which Garmadi (1981: 65) prefers the notion of the 'supra-norm'. The latter represents, in Garmadi's words, 'l'idéal esthétique ou socioculturel d'un milieu détenant prestige et autorité [...] l'existence de ce système d'instructions implique celle d'usages prohibés' ('the aesthetic or sociocultural ideal of a milieu with prestige and authority [...] the existence of this system of instructions implies that of prohibited usages'). Müller uses the term 'supernorm' to refer to a variety *outside* the core, in other words 'educated French', a writing-like language variety used in upper-group speech or found in literature and formal writing. The 'supernorm' has prestige and authority within a profession or a social group. Meillet (1921: 115) speaks of 'langues spéciales' (special languages). The languages of administration or science, as technical but still prestigious language varieties, stand midway between the core and the 'supernorm'. The low variety and the 'supernorm' constitute, to use Rey's (1972: 17) formulation, 'un pseudo-système [qui] se donne pour le système' ('a pseudo-system that asserts itself as the system'). Both of them could enjoy covert prestige with a particular group or social class. However, the 'supernorm', unlike the low variety, is to some extent 'codifié' ('codified') and 'normalisé' ('normalized') (Garmadi 1981: 56-57).



Graph 2: Some of the style labels in *Le Nouveau Petit Robert* (1993)

Graph 2, classifying some of the style labels in *Le Nouveau Petit Robert* (1993), shows that the three subgroups may interrelate. The labels *vieilli* (outdated) and *vx* (*vieux*, obsolete) can for example combine with *fam.* (*familier*, colloquial) or *pop.* (*populaire*, popular) as well as *unlabelled* or *lit.* (*littéraire*, literary) words. This is also the case with the label *néol.* (*néologisme*, neologism). Some unlabelled words can be rated as *péj.* (*péjoratif*, derogatory) or *inj.* (*injurieux*, offensive). Standard items can also have a derogatory or offensive connotation and receive the labels *péj.* or *inj.* Between the core and the low category, words are found labelled as *mod.* (*moderne*, currently used), *cour.* (*courant*, commonly used), *rég.* (*régional*, regional), *enf.* (*enfantin*, children's usage), *triv.* (*trivial*, coarse), *abusot* (*abusivement*, abusively used) and *imp.* (*impropre*, incorrect). Purists consider Anglicisms, labelled *ang.* (*anglicisme*) in the low category, as non-standard forms. While some have merged into French, they are no longer labelled as Anglicisms. Items labelled *vulg.* (*vulgaire*, vulgar), on the contrary, belong to low usage. As Lodge points out, the label *pop.* 'with its very nebulous sociological rather than stylistic basis [...] seems particularly inappropriate' (1989: 442-443), especially on account of the fact that social classes in France are not clearly definable. It could be said that both *fam.* and *pop.* are stylistic rather than social indicators on the low/high continuum. The latter variety, being more informal than the first, can be placed between *fam.* and *vulg.* on the graph. There is much disparity in dictionaries, even in *Le Petit Robert*. For no particular reason, some labels appear abbreviated, some not. In their judging of the different degrees of familiarity, dictionaries use labels such as *très fam.* (very familiar) or *très vulg.* (very vulgar). The label *arg.* (*argot*, slang) has been placed overlapping the low variety and the 'supernorm'. Although some of its usages are still traditionally regarded as having low status (dead slang), the 'langues de spécialité' (specialist languages, living slangs) belong to the 'supernorm'. The label *dial.* (*dialectal*, dialectal) is rated as a form of regional slang. Technical vocabularies associated with a particular occupation (professional slangs) (labelled *sport* (sports), *sc.* (*science*, science), *comm.* (*commercial*, marketing), *dr.* (*droit*, law), *industr.* (*industrie*, industry), *techn.* (*technique*, technical), *didact.* (*didactique*, didactic), and *off.* (*officiel*, official) by *Le Petit Robert*) likewise overlap with both the 'supernorm' and the high variety, some of these items being commonly used in spontaneous French. The terminology used by dictionaries is arguable, as words labelled *lit.*, *poétique* (poetical) and *arch.* (*architecture*, architecture) enjoy overt prestige, while all technical language varieties including slang have covert prestige among their users.

A comparison of *Le Larousse du XXème siècle* (1932) where only major style labels such as *fam.*, *pop.* and *arg.* were used with the more recent *Le Petit Larousse illustré* (1989) and *Le Nouveau Petit Robert* (1993) shows that the differentiation between style labels has become increasingly subtle (Abecassis 2004). The appearance of *cour.* (*courant*, commonly used) and *mod.* (*moderne*, currently used) gives a diachronic dimension to the diatopic, diaphasic or diastratic variations. The label *ang.* (*anglicisme*, Anglicism) has a negative connotation in

Le Robert but not in *Le Larousse*. Unlike *Le Larousse*, *Le Robert* uses a range of labels to refer to technical vocabularies (e.g. *sport* (sports), *sc.* (*science*, science), *comm.* (*commercial*, marketing), etc.).

As the differences between spoken and written forms become more blurred with the use of email and text messages, today's dictionaries aim at being more and more comprehensive with the inclusion of slang and technical jargons, as well as encyclopaedic knowledge. Rey's *Dictionnaire culturel en langue française* (2005) with more than 80 000 quotations and 65 000 entries presents words in their contexts with references to popular customs and beliefs and encapsulates a wide range of disciplines from mathematics to economics as well as music, history and psychology. Looking at monolingual dictionaries and their style labels made it possible to show how lexical entries are categorized, and the difficulty in finding consistent terminology that would encapsulate their lexicons. Nor is it possible for a dictionary to be exhaustive for practical reasons. Words are selected according to their usage, and archaisms tend to be omitted. In 2004 the French journalist Bernard Pivot, creator of the famous spelling championships, listed a hundred words which are threatened with becoming obsolete and which deserve to be saved by lexicographers. Dictionaries also have the function of preserving language usage which forms an integral part of a community's cultural heritage.

2. Types of dictionaries

If the *Oxford English Dictionary* excels as the English reference work par excellence, other types of dictionaries are used by learners of English. Second language learners tend to prefer bilingual dictionaries to monolingual ones as reference books (Martin 1998). However, bilingual dictionaries are often incompetently used with students only relying on a word-for-word translation or ignoring the whole range of translations a word can have. Monolingual dictionaries are often recommended by teachers of a second language, especially at the intermediate level, to encourage learners to think in the target language. Given the lack of research on this topic, it has been impossible to establish the superiority of monolingual dictionaries (Wingate 2002: 1). Not only are native speakers in search of a definition or the spelling of a word frequent users of general-purpose dictionaries, but language learners who are already proficient in the language and have a good grasp of the grammar and lexical basis, will also seek 'to extend their vocabulary into the more peripheral areas of the lexicon' (Leech and Nesi 1999: 300). Foreign language teaching generally focuses on the vitality of the language and tends to exclude obsolete forms¹ which the speaker is unlikely to encounter or neologisms on account of their non-standard nature. Apart from monolingual and bilingual dictionaries, there are bilingualized dictionaries which combine the function of both bilingual and monolingual dictionaries: 'Whereas bilingual dictionaries usually provide just an L1 synonym, bilingualized dictionaries include L2 definitions, L2 sentence

examples, as well as L1 synonyms. Bilingualized dictionaries were found to result in better comprehension of new words than either bilingual or monolingual dictionaries' (Hunt and Beglar 2005). The hybrid bilingualized learner's dictionary is a relatively new genre of dictionary which, although it has attracted some criticism, is thought to be suitable for all learners, whatever their proficiency, who prefer to rely on the source-language translation equivalents.

2.1 Language learner's dictionaries and lexical acquisition

The number of new editions of language learner's dictionaries released every year by Hachette, Oxford, Longman and Collins among others, mostly for English as a foreign language but also for French, German, Italian and Spanish, has increased over the past ten years, to the extent that it has been extremely difficult for learners to find the foreign language dictionary which will best suit their needs. There is no such thing as a perfect dictionary, just as there is no perfect language learner, and every user has different needs according to his/her culture, background, age or level.² The major hindrance in using a dictionary is that the user often does not possess the necessary skills to utilize it efficiently (Hartmann and James 1998) or is insufficiently guided by the preface to exploit its possibilities. Most students in British universities not only lack the dictionary-using abilities and the look-up strategies required to fully exploit on their own the resources of an English learner's dictionary, but they also often prefer bilingual to monolingual dictionaries (Nesi 2000).

Cowie (2000) has described in detail the early history of EFL dictionaries and their evolution since the early 1980s thanks to the input of large-scale corpora of English, the British National Corpus and the Bank of English,³ and their improvement to meet the needs of users, by employing authentic language, in this way becoming more user-friendly. While Johnson's dictionary (1755) drew extensively on literary quotations (Krishnamurthy 2002), vast corpora which comprise literary texts, newspaper articles and informal natural speech have yielded a wide variety of genuine, up-to-date examples which carry greater authority and are particularly enlightening in establishing the state of a language at a particular time. The history of EFL dictionaries started remotely from Europe in Japan and India, where the three pioneers of EFL dictionaries, H.E. Palmer, A.S. Hornby and M.P. West, taught and conducted extensive research in L2 teaching on both Japanese and Indian EFL students. At a first stage, Palmer (1933) and later West and Endicott (1935) as well as Palmer and Hornby (1937) started elaborating word-lists which Cowie (1999) refers to as the 'structured lexicon', that part of the lexis which, in everyday conversation, is the most frequently used. By focusing on this core vocabulary and using particular arrangements (each entry would include its range of meanings and derivatives), they encouraged the user to encode. West and Endicott's *New Method English Dictionary* effected the innovation that the vocabulary used in the definitions of each lexical item was kept to the very simplest. It is thought that a minimum of 2 000 words is necessary to understand a word definition in

a second language (Nation 2001: 292). Dictionaries had in the past resorted to mere synonyms or used complex vocabulary in their definitions of particular items which was a major impediment to students. Also in his controlled vocabulary West included a list of some of the most frequent prefixes and suffixes which users might encounter, thus giving them the foundation for word construction.

The earlier EFL dictionaries, in the tradition of general dictionaries, were essentially prescriptive, drew quotations from literary sources and prescribed colloquial examples (Krishnamurthy 2002). From the 1970s, EFL dictionaries endeavoured to be more descriptive and innovative by using IPA transcriptions for pronunciation, descriptive definitions, usage notes on various aspects of the target culture, institutions and customs which in an often very entertaining way inform the user on a wide range of topics, though sometimes verging too much on social stereotypes or academic terminology, and, in the case of Longman, phrasal verbs, information on collocations and grammar patterns, with specifications for the syntactic use of words. In addition, numerous pictures and illustrations⁴ clarify abstruse meanings or technical vocabulary and definitions favour invented examples which contextualized lexical items in authentic quotations. Dictionaries are now increasingly supplemented by practical exercises with explanations and corrections. There are generally three types of dictionaries most specifically targeting language learners of English such as the series edited by Collins COBUILD (see Table 1): student's dictionaries, learner's dictionaries and advanced learner's dictionaries.

Title	Level targeted	Capacity
<i>Collins COBUILD Advanced Learner's Dictionary</i>	Intermediate/ Advanced	1 167 pages 75 000 references 105 000 examples
<i>Collins COBUILD Concise Learner's Dictionary</i>	Intermediate	1 344 pages 60 000 references 55 000 examples
<i>Collins COBUILD Student's Dictionary</i>	Intermediate	1 088 pages 200 pages of English grammar

Table 1: Collins COBUILD Language Learner's Dictionaries

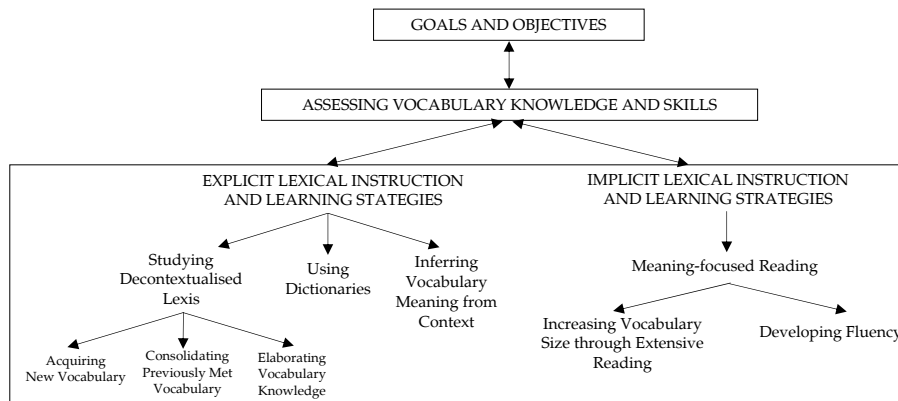
The less advanced the users targeted, the smaller the size of the lexicon and the fewer the references and examples, as can be seen when the statistics of the three Collins COBUILD dictionaries are compared. For a language learner's dictionary in contradistinction to any other English dictionary lexicographers have selected what constitutes the essential core of a language (Stein 2002: 35) and have opted in their definitions for full grammatical sentences with a simpler and more transparent vocabulary. Each definition shows grammatical patterns with additional sections on phrasal verbs and loan words, and word

usages are illustrated by several examples. The third edition of the *Collins COBUILD English Dictionary for Advanced Learners* includes a practical guide to dictionary use, and has an extensive coverage of American English. Dictionaries often focus on British English and lack references to dialects and regional varieties. Little can be found on regional varieties of English such as that spoken in the West Indies, New Zealand or Australia, and this is an aspect that requires improvement (Leech and Nesi 1999: 300). Besides, as Stein (2002: 75) notes, the latest EFL dictionaries are biased in favouring Received Pronunciation (RP) over American English. Each new edition of a dictionary intends to innovate with a newly devised user-friendly layout which would guide learners more quickly to the word sought. Special labelling and use of colour highlight the most frequently used lexical items. Coverage of today's 'trendy' English such as the language of computing and telecommunication is emphasized, which makes the volume more exhaustive (the dictionary accounts for 90% of written and spoken language). In the *MacMillan English Dictionary for Advanced Learners* (2002), red stars are used to tag the most frequently used items which constitute the core of the English language.

Dictionaries in electronic form also require from the students a high degree of practice and a careful study of the help menu. Navigational and searching skills will become advanced once users have familiarized themselves with the software. Computerized English dictionaries can contain over a million lexical items and phrases and also include access to multiple search functions in other languages. With new developments in computerization, CD-ROMs, and electronic and pocket dictionaries have been released to supplement paper editions of dictionaries. The practicality of these materials immediately comes to mind. Compared to paper editions, they are easy to carry, light and cheap, and the search for words is often faster (Weschler 2000). The use of multimedia has proven to help vocabulary acquisition (Chapelle 2001), but as to whether words learnt electronically by students are in the long run better memorized needs to be further investigated by lexicographers. Since the searching process has been accelerated by e-dictionaries and therefore exposure to the words is shorter, it may be doubted that memorization is as efficient. CD-ROMs provide the user with a considerable number of multimedia possibilities that can be exploited (game-like audio and video support, interactive exercises with different levels to choose from, hypertext links and internet links). Some are equipped with spelling games, exercises, lists of synonyms, antonyms and thesauri. The pronunciation, intonation of isolated lexical items or full sentences can now be heard, which for language learners is an invaluable resource. In addition, the user can become acquainted with pragmatic features particular to the target language such as body language and other gestures. Corpora are now available on-line, which has given the opportunity for both teachers and students to have direct and easy access to authentic data. Furthermore, concordance software like Scott's WordSmith Tools makes it pos-

sible for students and teachers alike to compile their own corpus and look for particular collocations or idioms.

The process of learning vocabulary and the problems children face in the learning process have been extensively studied by Siskind (1997); similarly the mental aspects of how the mind organizes and processes vocabulary has been investigated (Aitchison 1987). It could be agreed with Aarts (1999: 16) that ideal learner's dictionaries should perform two major roles: 'they should enable students to decode what they do not understand and, at the same time, serve as instruments that enable them to produce their own texts'. The dictionary is indeed an instrument in language learners' long process of building up their own lexicons and developing fluency in another language. As they come across new vocabulary, they also learn strategies, and reusing them in context will improve retention. Not using the acquired vocabulary or lack of practice in context, on the contrary, will inevitably lead to vocabulary loss. Since, for most language learners working in their country of origin, exposure to a natural speaking environment is missing, they must have recourse to dictionaries and written sources. Language acquisition may be intentional with 'the learner's deliberate decision to commit information to memory' (Laufer and Hulstijn 2001: 1) (explicit learning), or incidental with the learner broadening his/her vocabulary knowledge and developing fluency without being aware of it (implicit learning). Graph 3 extracted from Hunt and Beglar (2005) shows the framework of language acquisition.



Graph 3: Framework for developing vocabulary knowledge and skills (Hunt and Beglar 2005)

As shown in this graph, the use of dictionaries plays an essential part in lexical acquisition, with the studying of decontextualized lexis and the inferring of meaning from the context. When engaging in extensive reading, the EFL learner comes across words that he/she has learnt in context, and, with implicit exposure to idioms, collocations, discourse markers and syntactic structures, integrates them gradually into long-term memory. Much progress has been

achieved by dictionary makers to improve on the presentation of lexical items and give 'today's user of English unrivalled access to the English they need' (CIDE). Comparing different learner's dictionaries to assess their suitability, such as statistically quantifying their coverage of verb, adjective and noun patterns, Klotz (1999: 42) has shown that one was not 'clearly better than the others', but that all four dictionaries he studied (CIDE, COBUILD2, LDOCE3 and OALD5) gave more emphasis to verb patterns and accorded less detail to the treatment of adjective and noun patterns. Most researchers agree that the learner's dictionary of today bears no comparison with previous ones because of the considerable input of corpus linguistics, but that there is still much to be done to improve its content and accessibility and that more research on its appropriateness is required. Dictionary makers, in their quest for the perfect learning instrument, both easy to understand and to use, keep on improving by benefiting from lexicographers' constructive remarks and learners' expectations and motivations, but the dictionary is only one part of lexical acquisition and is very much dependent on other external factors such as the learners' abilities, their age, background and the learning environment.

The learner's dictionary of the future will always strive to be more comprehensive and more attractive to the user and, complying with the desiderata of lexicographers, teachers and students alike, will endeavour to present information in a pedagogic, clear and challenging way, whilst relying on authentic material. Will electronic monolingual and bilingual dictionaries whose unlimited potential can link up not only with other on-line dictionaries but also with computer tools, spell checkers and thesauri ever supplant printed dictionaries? Though users will agree that their speed is convenient, it has not been shown yet whether this new generation of dictionaries substantially improves language learners' acquisition of native-like competence and knowledge, and broadens, in the long run, their lexicon. As there is a whole range of levels from beginners to advanced learners who work more independently, and as there are different learners in different teaching environments, it is likely that a proposed model of dictionary effectiveness is to a large extent subjective and, if left to the learners' judgement, what suits the needs of one user might be felt inappropriate for another. Future contributions on this subject will make it possible to evaluate the effectiveness of e-dictionaries and answer the question of whether the theory of a perfect dictionary is not mainly ideological.

Endnotes

1. See Stein (2002: 216-229) for the treatment of archaisms by EFL dictionaries.
2. A large number of studies have been conducted on the users of dictionaries (cf., for example, Dolezal and McCreary 1999, Bogaards 2003).
3. The Bank of English which is used for Collins COBUILD editions now contains over 320 million words.
4. The inclusion of illustrations is not something new. Already in the sixteenth century, Thomas

Elyot's *Dictionary* (1538) was the first printed English dictionary to include illustrations, followed by Richard Huloet (1552), Thomas Cooper (1565), John Bullokar (1612) and Thomas Blount's illustrated word list in *Glossographia* (1656) (see Stein 2002: 172-173).

Bibliography

- Aarts, F. 1999. Syntactic Information in OALD5, LDOCE3, COBUILD2 and CIDE. Herbst, T. and K. Popp (Eds.). 1999: 15-32.
- Abecassis, M. 2004. Analyse diachronique du vocabulaire familier des films des années 30. *French Studies in Southern Africa* 33: 1-17.
- Aitchison, J. 1987. *Words in the Mind*. Oxford: Blackwell.
- Bogaards, P. 2003. Uses and Users of Dictionaries. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 26-30. Amsterdam/Philadelphia: John Benjamins.
- Caradec, F. 1977. *Dictionnaire du français argotique et populaire*. Paris: Larousse.
- Chapelle, C.A. 2001. *Teaching English as a Second or Foreign Language*. Boston: Heinle and Heinle.
- CIDE: *Cambridge International Dictionary of English*. (Edited by P. Procter.) 1995. Cambridge: Cambridge University Press.
- COBUILD: *Collins COBUILD English Dictionary for Advanced Learners*. (Edited by J. Sinclair 1987, 1995.) Sixth Revised Edition with CD-ROM. 2006. Glasgow: HarperCollins.
- COBUILD2: *Collins COBUILD English Dictionary*. (Edited by J. Sinclair.) Second Edition. 1995. London: HarperCollins.
- Cowie, A.P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: Clarendon Press.
- Cowie, A.P. 2000. The EFL Dictionary Pioneers and their Legacies [online]. *Kernerman Dictionary News* 8. <http://kdictionary.com/newsletter/kdn8-1.html>.
- Dolezal, F.T. and 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.
- Garmadi, J. 1981. *La sociolinguistique*. Paris: Presses Universitaires de France.
- Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Herbst T. and K. Popp (Eds.). 1999. *The Perfect Learners' Dictionary (?)*. Lexicographica. Series Maior 95. Tübingen: Max Niemeyer.
- Houdebine, A.-M. 1982. Norme et imaginaire linguistique. *Le français moderne* 50: 42-51.
- Hunt, A. and D. Beglar. 2005. A Framework for Developing EFL Reading Vocabulary [online]. *Reading in a Foreign Language* 17(1). <http://nflrc.hawaii.edu/rfl/April2005/hunt/hunt.html>.
- Imbs, P. 1969. Les niveaux de langue dans le dictionnaire. *Le français dans le monde* 69: 51-60.
- Johnson, S. 1755. *A Dictionary of the English Language*. London: Longman.
- Klotz, M. 1999. Word Complementation in English Learners' Dictionaries — A Quantitative Study of CIDE, COBUILD2, LDOCE3 and OALD5. Herbst T. and K. Popp (Eds.). 1999: 33-43.
- Krishnamurthy, R. 2002. The Corpus Revolution in EFL Dictionaries [online]. *Kernerman Dictionary News* 10. <http://kdictionary.com/newsletter/kdn10-9.html>.
- Laufer, B. and J. Hulstijn. 2001. Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-induced Involvement. *Applied Linguistics* 22: 1-26.
- LDOCE3: *Longman Dictionary of Contemporary English*. (Edited by D. Summers.) Third Edition. 1995. Harlow: Longman.

- Leech, G. and H. Nesi.** 1999. Moving Towards Perfection: The Learners' (Electronic) Dictionary of the Future. Herbst, T. and K. Popp (Eds.). 1999: 295-306.
- Le Larousse du XXème siècle.* 1932. Paris: Larousse.
- Le Nouveau Petit Robert.* 1993. Paris: Le Robert.
- Le Petit Larousse illustré.* 1989. Paris: Larousse.
- Lodge, R.A.** 1989. Speaker's Perception of Non-standard Vocabulary in French. *Zeitschrift für Romanische Philologie* 105 (5-6): 427-44.
- Macmillan English Dictionary for Advanced Learners.* (Edited by M. Rundell and G. Fox.) 2002. London: Macmillan.
- Martin, R.** 1998. One Benefit of Monolingual Dictionaries in the Writing Classroom [online]. *The Internet TESL Journal* 4(10). <http://iteslj.org/Articles/Martin-Dictionaries.html>.
- Merle, P.** 1986. *Dictionnaire du français branché.* Paris: Le Seuil.
- Michel, J.-F.** 1807. *Dictionnaire des expressions vicieuses usitées dans un grand nombre de départements et notamment dans la ci-devant Lorraine.* Paris: Bentoux.
- Milroy, J. and L. Milroy.** 1992. *Authority in Language: Investigating Language Prescription and Standardisation.* London/New York: Routledge.
- Müller, B.** 1985. *Le Français d'aujourd'hui.* Paris: Klincksieck.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.
- Nesi, H.** 2000. *The Use and Abuse of EFL Dictionaries: How Learners of English as a Foreign Language Read and Interpret Dictionary Entries.* Lexicographica. Series Maior 98. Tübingen: Max Niemeyer.
- OALD5: *Oxford Advanced Learner's Dictionary.* (Edited by J. Crowther.) Fifth Edition. 2005. Oxford: Oxford University Press.
- OALDCE: *Oxford Advanced Learner's Dictionary of Current English.* (Edited by S. Wehmeier and A.S. Hornby.) Seventh Edition with CD-ROM. 2005. Oxford: Oxford University Press. <http://www.oup.com/elt/catalogue/teachersites/oald7/?cc=gb>.
- Oxford English Dictionary.* (Edited by J.A.H. Murray.) 1884-1933, 1933. Oxford: Oxford University Press.
- Oudin, A.** 1640. *Curiositez françoises.* Paris: A. de Sommaville.
- Palmer, H.E.** 1933. *Second Interim Report on English Collocations.* Tokyo: Kaitakusha.
- Palmer, H.E. and A.S. Hornby.** 1937. *Thousand-Word English.* London: George Harrap.
- Pivot, B.** 2004. *100 mots à sauver.* Paris: Albin Michel.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik.** 1985. *A Comprehensive Grammar of the English Language.* London/New York: Longman.
- Rey, A.** 1972. Usages, jugements et prescriptions linguistiques. *Langue française* 16: 4-28.
- Rey, A.** 2005. *Dictionnaire culturel en langue française.* Paris: Le Robert.
- Rigaud, L.** 1878. *Dictionnaire du jargon parisien.* Paris: P. Ollendorf.
- Scott, M.** 2004. *WordSmith Tools.* Version 4. Oxford: Oxford University Press.
- Shaw, J.** 1997. *Contributions to a Study of the Printed Dictionary in France before 1539* [online]. <http://www.chass.utoronto.ca/~wulftric/edicta/shaw/frame.htm>.
- Siskind, J.** 1997. A Computation Study of Cross-situational Techniques. Brent. M.R. (Ed.). 1997. *Computational Approaches to Language Acquisition:* 39-93. Amsterdam: Elsevier Science Publisher.

- Stein, G.** 2002. *Better Words: Evaluating EFL Dictionaries*. Exeter: University of Exeter Press.
- Webster's New Collegiate Dictionary*. (Edited by H.B. Woolf.). 1976. Springfield MA: Merriam.
- 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>.
- West, M.P. and J.G. Endicott.** 1935. *The New Method English Dictionary*. London: Longmans, Green.
- Wingate, U.** 2002. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learner of German*. Lexicographica. Series Maior 112. Tübingen: Max Niemeyer.
- Wooldridge, T.R.** 1977. *Les débuts de la lexicographie française. Estienne, Nicot et le Thresor de la langue françoise (1606)*. Toronto: University of Toronto Press.

The Effectiveness of Dictionary Examples in Decoding: The Case of Kuwaiti Learners of English*

Hashan Al-Ajmi, *Department of English, Faculty of Arts,
Kuwait University, Kuwait (hashan98@yahoo.com)*

Abstract: This study tries to shed light on the role of dictionary examples in the comprehension of word meanings. An experimental procedure has been devised whereby two groups of students with English as major subject at Kuwait University were asked to provide the Arabic equivalents for ten English headwords. The first group was given a list of entries for these words copied from the *Oxford Advanced Learner's Dictionary* (OALD) while the second group had to read the same list but without illustrative examples. Results indicate that the students' decoding performance was negatively affected by the presence of illustrative examples in the dictionary entry.

Keywords: ARABIC, BILINGUAL DICTIONARY, COMPREHENSION, EFL DICTIONARY, ILLUSTRATIVE EXAMPLE, MONOLINGUAL DICTIONARY, TRANSLATION

Opsomming: Die doeltreffendheid van woordeboekvoorbeelde by dekodeerding: Die geval van Koeweiti-aanleerders van Engels. Hierdie studie probeer lig werp op die rol van woordeboekvoorbeelde by die verstaan van woordbetekenisse. 'n Eksperimentele metode is ontwerp waarby twee groepe studente met Engels as hoofvak by die Universiteit van Koeweit gevra is om Arabiese ekwivalente vir tien Engelse trefwoorde te verskaf. Aan die eerste groep is 'n lys inskrywings van hierdie woorde oorgeneem uit die *Oxford Advanced Learner's Dictionary* (OALD) gegee, terwyl die tweede groep dieselfde lys moes lees, maar sonder verduidelikende voorbeelde. Resultate het aangedui dat die studente se dekodeerende prestasie negatief beïnvloed is deur die teenwoordigheid van verduidelikende voorbeelde in die woordeboekinskrywing.

Sleutelwoorde: ARABIES, TWEETALIGE WOORDEBOEK, BEGRIP, EVT-WOORDEBOEK, VERDUIDELIKENDE VOORBEELD, EENTALIGE WOORDEBOEK, VERTALING

Introduction

Verbal illustrations have always been considered an essential component of dictionary entries and a criterion by which metalexigraphers and reviewers evaluate dictionaries. However, as shown in various lexicographic traditions,

* This article is a revised version of a paper 'The Effectiveness of EFL Dictionary Examples in Kuwaiti Learners' Comprehension of Entry Words' presented at the Thirteenth International Symposium on Lexicography, University of Copenhagen, Copenhagen, Denmark, 3-5 May 2007.

illustrative examples have no clearly defined or agreed-upon functions. In the past, exemplification was utilized in Arabic dictionaries in the form of quotations aimed at proving the existence of rare words in the language. Arab lexicographers took these illustrative examples mainly from the Quran, Hadith and poetry (Al-Kasimi 1977). In the lexicographic heritage of English, Johnson was the first to use examples systematically to illustrate the words defined in the entries. His examples were taken from the writings of great authors like Shakespeare, Milton, Dryden and Addison (Cowie 1990). The appearance of EFL dictionaries such as the *Oxford Advanced Learner's Dictionary* (OALD) in 1947 has marked a new trend in which dictionary examples were revised and increased to perform an essential function, i.e. to illustrate grammatical patterns and throw light on the meaning of the entry (Hornby 1965). A detailed description classifying the functions of illustrative examples into active and passive functions has been provided by Drysdale (1987: 218): to supplement information in a definition, to show the entry word in context, to distinguish one meaning from another, to illustrate grammatical patterns, to show other typical collocations and to indicate appropriate registers or stylistic levels.

Thanks to the information revolution, the electronic corpus has become the main source of dictionary examples, and one EFL dictionary, the *Collins COBUILD English Language Dictionary*, relies entirely on a corpus for its examples (cf. Sinclair 1987). Yet some lexicographers still prefer to invent their examples. Cowie (1989: 62) stresses 'the role of the dictionary compiler as the deliberate controller and manipulator of illustrative material' and defends the traditional invented examples because they can be shaped in the learners' interests. He also believes that the extensive use of examples is an alternative (or necessary addition) to clearer labelling as a means of indicating syntactic patterns and advises that examples taken from actually occurring texts should not be accorded a sacred status (Cowie 1989: 45). Also, Minaeva (1992) argues for examples created by lexicographers rather than examples taken out of context from a large corpus while Mitmann (1995) criticizes the corpus examples of COBUILD because of their length and complexity. On the other hand, there are those who regard a corpus as a reliable source of illustrative material. Indeed, all modern EFL dictionaries depend to different degrees on corpus data in every aspect of their text, so the differences now lie in the degree to which corpus material is processed on its way to the examples (Rundell 1998: 334). Criticism is often voiced against invented examples by corpus proponents like Sinclair (1984: 4) who states that 'when examples are concocted by the same lexicographer, they have no value at all. It can be claimed that they illustrate the word in use, but one thing we do know is that usage cannot be thought up — it can only occur'. Atkins (1991) illustrates how concordances of a corpus are useful in constructing fuller entries with a finer breakdown of senses. As to the users, Humblé (1998) argues that advanced users can benefit from authentic corpus examples, while intermediate learners can learn from 'controlled' exam-

ples, and Potter (1998) indicates that carefully chosen corpus examples are more helpful for users of English learners' dictionaries than invented examples.

However, users in different parts of the world still face serious difficulties trying to extract dictionary explanatory and illustrative data in both the decoding and encoding modes. For example, Kharma (1984) observes that most students at Kuwait University use a bilingual dictionary all the time and that these students seem to have difficulty in understanding the definitions used in their English monolingual dictionaries. This observation is shared by El-Badry (1990) in whose study of Arab students' dictionary use the subjects said that they find monolingual dictionaries difficult to use because their definitions are ambiguous and their examples insufficient. In Hong Kong, Chi (1998) also found that most university students own a bilingual English–Chinese dictionary for comprehension, but that very few students own monolingual English learners' dictionaries, a possible indication of their difficulties with them. But bilingual dictionaries may not provide the needed information. In the study by Marengo (1987: 236) of Italian students' views of their bilingual dictionaries, results of the questionnaire show that examples are not sufficiently informative.

The availability of examples in the dictionary entry does not mean that users would be able to utilize the information. Only those who know how to read them can achieve that objective (Maingay and Rundell 1987). It is impossible for a dictionary to provide all possible constructions, collocations and contexts of usage of a given word. Another inherent shortcoming of exemplification is outlined by Whitcut (1985: 77): 'Examples can show how a word can be used, but not how it can't.' A detailed account of the problems associated with dictionary examples and definitions is given by Creamer (1987). These problems include: an incomplete definition in the entry combined with a poor example, a lack of sense discrimination in the example, inappropriate placement of the example in the entry, and the absence of examples. He refers to problems with definitions indicating that definitions written too broadly or too narrowly may confuse users.

Entries combining definitions with examples may affect users' response to the data of the entry. In their experiment, Cumming et al. (1994) found that students preferred COBUILD's sentence definitions plus a usage example (71%) over the phrasal definitions plus examples (27%) in the *Longman Dictionary of Contemporary English* (LDOCE). And Herbst (1986) found that users had more success in comprehending entries constructed with a controlled vocabulary in comparison to the non-controlled dictionary. But difficulties with monolingual definitions are not restricted to foreign learners of English. McKeown (1991) found that young native speakers do not learn much from the traditional dictionary format and that revised definitions yielded better results.

The issue of illustrative examples in English lexicography is summarized by Jackson (2002: 181): 'Examples are seen as playing a crucial role in EFL dictionaries. A number of questions need to be asked by the critic, relating to: the

extent of the use of examples, what role they are seen to play in exemplification, where the examples come from (corpus or invented), and how consistently the dictionary's policy on examples is implemented.'

Previous studies on examples

A few studies dealt with the issue of the presence versus absence of examples in the dictionary entry as well as the issue of whether examples should be invented or derived from a corpus. As for corpus derived examples, Laufer (1992) conducted a test to determine the effectiveness of the two types in the comprehension and use of new vocabulary items. Her results indicated that invented examples were more effective than COBUILD's corpus examples. And in 1993 Laufer performed another experiment to discover the effect of dictionary definitions and examples on the use, comprehension and translation of new L2 words by 43 first-year students at the Hebrew University. Some of the subjects were given definitions, another group were given examples while the third group dealt with combined entries, i.e. both a definition and an example. Results of her study showed that a combined entry yields the best results in comprehension and that a definition alone is better than only an example. Laufer (1993: 139) found that 'the understanding of new words improves more when a definition is added than when an example is added ... in production, unlike in comprehension, an example alone will be just as efficient as a definition alone'.

In another study by Nesi (1996), 40 non-native speakers of English studying in Britain were tested to discover how much help examples can be in the productive use in sentences of 18 unfamiliar words. Two versions were prepared taking entries from LDOCE. In the first version, examples were removed from the first half of the entries and from the second half of the second version. Nesi (1996: 201) found that 'there was ... no statistical evidence that subjects were more productively competent when they were provided with illustrative examples'. A more recent study (Al-Bader 2007) tested Kuwait University students using four EFL dictionaries to verify the effectiveness of definitions only versus definitions plus examples. However, Al-Bader (2007: 101) found that 'results were not enough to decide whether the existence of sentences along with definitions play a substantial role in understanding the meaning of English words, as most of the students have identified the meaning of the words with or without the assistance of illustrative examples'.

It is noticeable that previous studies put more focus on the role of examples in production and less on comprehension. But it is known from several user studies (e.g. Battenburg 1991, Al-Ajmi 1992, etc.) that both types of dictionary are mostly used for decoding rather than encoding linguistic activities (also see Cowie 1999).

The study

Given the discrepancy in the results of previous studies, the present investigation aims to test the effectiveness of illustrative examples when using the EFL dictionary to understand unfamiliar words. Hence, the focus will be on reading comprehension since it is the linguistic activity in which learners often refer to their dictionaries. The study tries to find answers to the following research questions:

- (1) Is the presence of an example better than its absence in the EFL dictionary when used for comprehension?
- (2) Are there any specific patterns followed by students as shown in their translations?
- (3) How close are their translations to bilingual dictionary equivalents?

Ten words were selected and pilot-tested to ascertain their difficulty. They are: *aftermath, aisle, bona fide, deflect, elide, grandeur, lingua franca, ominous, repulsive, restive*. The pilot test has shown that the majority of respondents were in need of consulting their dictionaries for the test items either for total lack of knowledge or uncertainty about their meanings. Items that were to be looked up as a result of uncertainty (i.e. the student wanted to double check) were included in the list as this reflects a real situation of dictionary use. Dictionaries are sometimes consulted even when the meaning of a word can be guessed. Therefore, all the test items are expected to be looked up by the majority in real-life dictionary consultation. A list containing the entries of these words was taken from the OALD and two versions were prepared, one with definitions plus examples (see Appendix I) and the other with definitions only (see Appendix II).

The subjects of this study were 54 juniors at the Department of English, Faculty of Arts, Kuwait University enrolled in translation courses as well as English writing and conversation courses, all of which require extensive use of both bilingual and monolingual English dictionaries. Half of them were asked to translate the headwords after carefully reading the relevant definitions and examples. The other half of the subjects were asked to read the definitions only, then translate the headwords. Translation here is used as a tool for measuring how well the words were understood by this linguistically homogeneous group. It is believed that this method may yield more accurate results than composition of sentences containing the test words. The latter method may lead students to produce sentences that simulate the examples or definitions and do not guarantee that it will be known for certain whether the subjects understood the meanings of the entry words. Moreover, when writing, a student may adopt some strategies that conceal his/her inability to understand word meanings, and examples, according to Nesi (1996: 203), do not prohibit the use of alternative structures. It is not possible in this study which deals with

one dictionary title to determine whether examples derived from a corpus excel over those invented by lexicographers or vice versa. Also, the study does not concern itself with revealing user preferences with regard to types of examples as in the investigation by Cumming et al. (1994).

Correct and incorrect responses were determined by comparing students' translations with Arabic equivalents in the English–Arabic dictionary Al-Mawrid, the most popular bilingual dictionary in the Arab world. Accordingly, responses were analyzed, then classified into the following:

- (1) The translation is correct and identical with Al-Mawrid
- (2) The translation is correct and synonymous with Al-Mawrid
- (3) No translation has been given
- (4) The wrong translation is given
- (5) The translation is a derivative, but related to Al-Mawrid
- (6) The translation is a derivative, but unrelated to Al-Mawrid

Thereafter comparisons were made between translations given by those who read definitions only and translations by those who used definitions plus examples. Table 1 below shows the classification of the results of the two groups:

Table 1: Categories of student responses

Category	Definition only (n = 270)	%	Definition + Example (n = 270)	%
The translation is correct and identical with Al-Mawrid	34	12.6	20	7.4
The translation is correct and synonymous with Al-Mawrid	68	25.2	68	25.2
No translation has been given	23	8.5	16	5.9
The wrong translation is given	97	35.9	123	45.6
The translation is a derivative, but related to Al-Mawrid	13	4.8	10	3.7
The translation is a derivative, but unrelated to Al-Mawrid	35	13.0	33	12.2

The above categories are used as indicators of the students' ability to understand information in the dictionary entry (categories 1, 2 and 5) or their inability to do so (categories 3, 4 and 6).

Results and discussion

The results show a significant discrepancy between the two sets of wrong translations, i.e. by those using definitions only and those using definitions plus examples. Students translating the headwords with the help of definitions plus examples made more errors than those who relied on definitions only. Out of the 327 instances in categories 3, 4, and 6 that indicate inability to comprehend the entry 155 (47.4%) were associated with the use of definitions only while 172 (52.6%) were recorded for those who used definitions plus examples. There were no statistically significant differences between the two groups with regard to the derivationally inaccurate translations. Therefore, it can be seen that students' ability to comprehend words was less when definitions were combined with examples.

When explaining these results which indicate the ineffectiveness of examples the EFL dictionary itself may have to be addressed. This dictionary type is supposed to depart from the definitional style of native-speaker dictionaries, i.e. by providing simplified definitions that contain high-frequency words or controlled vocabulary. It seems that this objective has been achieved and that most problems still relate to the examples (cf. Creamer 1987). The first function of an example as described by Drysdale (1987), to supplement information in a definition, has not been realized in a way that facilitates users' success in comprehending word meanings. It appears that the policy followed in the construction of definitions in the OALD excels over that adopted in its examples. The language of its definitions might be easier and clearer than the language of the examples which might contain other unfamiliar words, thus leading to confusion and more text to be processed by the definition-plus-example group. This is shown in the large number of student equivalents that translate some words in the examples or definitions instead of the entry words. Perhaps these students thought that the example contains synonyms of the entry word and decided to translate these familiar words in the example. This tendency to focus on easy words in a definition or example has also resulted in translations of words with different parts of speech from the headwords. For example, the headword *repulsive* was translated by some students with the Arabic noun *kurh*, a direct equivalent of *dislike* in the definition, an indicator of the students' habit of not reading the whole text of the entry but stopping at the point where they think the needed information has been found. This occurred despite the presence of examples showing that *repulsive* is an adjective: *a repulsive sight/smell/person/habit*. This pattern may be explained with reference to the vocabulary and structure of definitions as well. Easy words in some definitions belong to a different part of speech as in the entry for the word *grandeur*. Some students translated the word *great* in the definition 'the quality of being great' into its corresponding Arabic adjective. It seems that the use of easier English synonyms of the headword would here have yielded better results owing to better understanding of the entry word. Also, it was noticed that students tend

to concentrate on the first easy word in the definition or example even when there is more than one sense. For example, the term *bona fide* was given the Arabic translation of 'legal', the third sense in its entry after 'genuine' and 'without deception'.

These results may also be explained by the fact that these learners are accustomed to the use of the bilingual dictionary where quick answers in the form of translation equivalents in the native language can be found. The test items (definitions plus examples) confronted the subjects with longer answers in the foreign language thus leading to more difficulties with the dictionary text.

The results show a low number of correct translations identical to Al-Mawrid equivalents in the case of definitions only (12%) against 7.4% for definitions plus examples. A possible explanation is that a shorter entry, without examples, means less distraction, thus enabling the subjects better to concentrate on the general concept of the word in Arabic. However, these translations were fewer than those that were correct but not listed in Al-Mawrid. The policy followed in this bilingual dictionary of providing a few Arabic synonyms that often do not agree with students' expectations should therefore be revised. The same state of affairs is noticed with translations presented by students, as most of their responses in this category were not related to the dictionary equivalents. This may also be attributed to the diglossic situation of Arabic which results in differences between the lexical repertoire employed in ideal written forms and the mental lexicon shared by Arab students.

Conclusions

This study has found that the provision of examples along with definitions negatively affects students' ability to understand unfamiliar English words. This finding clearly contradicts the common belief that examples are useful in both comprehension and production. It should, nevertheless, lead dictionary makers to think seriously about solutions to problems of constructing examples and definitions. EFL learners may be in need of easier examples and new ways for the construction of definitions, taking into account the problems faced by these learners when sorting through dictionary information (cf. Al-Ajmi 2002). Corpus data coupled with research-based knowledge of EFL learners' needs can be utilized in compiling more user-friendly dictionaries. A combination of both corpus examples and invented examples seems to be the viable approach that guarantees the provision of natural and usable examples. On the other hand, students need to be taught how to use their EFL dictionaries effectively and made aware of the information systems therein. As for studies on the role of examples in comprehension, researchers may obtain more accurate results if they opt for other research methods such as think-aloud protocols in order to understand what goes on in the users' minds while reading a dictionary entry.

References

- Al-Ajmi, H.** 1992. *The Use of Monolingual and Bilingual Arabic–English Dictionaries in Kuwait: An Experimental Investigation into the Dictionaries Used and Reference Skills Deployed by University Students of Arts and Science*. Unpublished Ph.D. Thesis. Leeds: University of Leeds.
- Al-Ajmi, H.** 2002. Which Microstructural Features of Bilingual Dictionaries Affect Users' Look-Up Performance? *International Journal of Lexicography* 15(2): 119-131.
- Al-Bader, Y.** 2007. *The Provision of Illustrative Examples in English Advanced Learners' Dictionaries*. Unpublished M.A. Thesis. Bangor: University of Wales.
- Al-Kasimi, A.M.** 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Atkins, B.T.** 1991. Corpus Lexicography: The Bilingual Dimension. Zampolli, A. et al. (Eds.). 1991. *Computational Lexicology and Lexicography. Vol. I*: 43-64. Pisa: Giardini Editori e Stampatori.
- Atkins, B.T. and K. Varantola.** 1998. Monitoring Dictionary Use. Atkins, B.T. (Ed.). *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*: 83-122. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Battenburg, J.** 1991. *English Monolingual Learner's Dictionaries: A User-Oriented Study*. Lexicographica. Series Maior 39. Tübingen: Max Niemeyer.
- Chi, M.** 1998. Teaching Dictionary Skills in the Classroom. Fontenelle, T. et al. (Eds.). 1998: 565-577..
- Cowie, A.P.** 1989. The Language of Examples in English Learners' Dictionaries. James, G. (Ed.). 1989. *Lexicographers and their Works*: 55-65. Exeter Linguistic Studies 14. Exeter: University of Exeter.
- Cowie, A.P.** 1990. Language as Words: Lexicography. Collinge, N.E. (Ed.). 1990. *An Encyclopedia of Language*: 671-700. London/New York: Routledge.
- Cowie, A.P.** 1999. *English Dictionaries for Foreign Learners. A History*. Oxford: Clarendon Press.
- Cowie, A.P. (Ed.)**. 1987. *The Dictionary and the Language Learner. Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*. Lexicographica. Series Maior 17. Tübingen: Max Niemeyer.
- Creamer, T.** 1987. Beyond the Definition: Some Problems with Examples in Recent Chinese–English and English–Chinese Bilingual Dictionaries. Cowie, A.P. (Ed.). 1987: 238-245.
- Crowther, J. (Ed.)**. 1995. *Oxford Advanced Learner's Dictionary of Current English*. Fifth Edition. Oxford: Oxford University Press.
- Cumming, G. et al.** 1994. On-line Lexical Resources for Language Learners: Assessment of Some Approaches to Word Definition. *System* 22(3): 369-377.
- Drysdale, P.D.** 1987. The Role of Examples in a Learner's Dictionary. Cowie, A.P. (Ed.). 1987: 213-223.
- El-Badry, N.** 1990. Arabic Lexicography in Northern Africa, with Special Reference to Egypt. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa*: 36-43. Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Fontenelle, T. et al. (Eds.)**. 1998. *Actes EURALEX '98 Proceedings. Papers Submitted to the Eighth EURALEX International Congress on Lexicography in Liège, Belgium*: 565-577. Liège: English and Dutch Departments, University of Liège.
- Hartmann, R.R.K.** 1983. The Bilingual Learner's Dictionary and its Uses. *Multilingua* 2(4): 195-201.
- Hartmann, R.R.K. (Ed.)**. 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.

- Herbst, T.** 1986. Defining with a Controlled Defining Vocabulary in Foreign Learners' Dictionaries. *Lexicographica* 2: 101-119.
- Hornby, A.S.** 1965. Some Problems of Lexicography. *English Language Teaching* 19(3): 104-110.
- Humblié, P.** 1998. The Use of Authentic, Made-up, and 'Controlled' Examples in Foreign Language Dictionaries. Fontenelle, T. et al. (Eds.). 1998: 593-599.
- Humblié, P.** 2001. *Dictionaries and Language Learners*. Frankfurt: Haag und Herchen.
- Jackson, H.** 2002. *Lexicography. An Introduction*. London/New York: Routledge.
- Kharma, N.** 1984. Contextualization and the Bilingual Learner's Dictionary. Hartmann, R.R.K. (Ed.). 1984: 199-206.
- Lauffer, B.** 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. Tommola, H. et al. (Eds.). 1992: 71-76.
- Lauffer, 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.
- Maingay, S. and M. Rundell.** 1987. Anticipating Learners' Errors — Implications for Dictionary Writers. Cowie, A.P. (Ed.). 1987: 128-135.
- Marello, C.** 1987. Examples in Contemporary Italian Bilingual Dictionaries. Cowie, A.P. (Ed.). 1987: 224-237.
- McKeown, M.** 1991. Learning Word Meanings from Definitions. Problems and Potential. Schwanenflugel, P. (Ed.). 1991. *The Psychology of Word Meanings*: 137-156. Hillsdale, NJ: Lawrence Erlbaum.
- Minaeva, L.** 1992. Dictionary Examples: Friends or Foes? Tommola, H. et al. (Eds.). 1992: 77-80.
- Mitmann, B.** 1995. *Examples in English Monolingual Dictionaries*. Unpublished M.A. Thesis. Erlangen-Nürnberg: Friedrich-Alexander University.
- Nesi, H.** 1996. The Role of Illustrative Examples in Productive Dictionary Use. *Dictionaries* 17: 198-206.
- Potter, L.** 1998. Setting a Good Example: What Kind of Examples Best Serve the Users of Learners' Dictionaries? Fontenelle, T. et al. (Eds.). 1998: 357-362.
- Rundell, M.** 1998. Recent Trends in English Pedagogical Lexicography. *International Journal of Lexicography* 11 (4): 314-342.
- Sinclair, J.** 1984. Lexicography as an Academic Subject. Hartmann, R.R.K. (Ed.). 1984: 3-12.
- Sinclair, J. (Ed.).** 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins.
- Tommola, H. et al. (Eds.).** 1992. *EURALEX '92 Proceedings. Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland, 4-9 August 1992*. Studia Translologica. Serie A(1-2). Tampere: Department of Translation Studies, University of Tampere.
- Whitcut, J.** 1985. Usage Notes in Dictionaries: The Needs of the Learner and the Native Speaker. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 75-80. ELT Documents 120. Oxford: Pergamon Press.

Appendix I: List of test entries (definition plus example)

1. **aftermath** *n* (*usu sing*) the circumstances that follow and are a consequence of an event, esp of an unpleasant one: *the rebuilding which took place in the aftermath of the war.*
2. **aisle** *n* a passage between rows of seats in a church, theatre, railway carriage, etc or between rows of shelves in a shop: *The bride and groom walked slowly down the aisle (ie after their wedding ceremony) ◦ An aisle seat (ie one next to the passage in an aircraft etc).*
3. **bona fide** *adj* [esp attrib], *adv* genuine(ly); without deception; legal(ly): *a bona fide agreement / contract / deal.*
4. **deflect** *v* change or make sth change direction, esp after hitting sth: *The bullet deflected from the wall ◦ One of the defenders accidentally deflected the ball into the net ◦ The chairman tried to deflect the shareholders' criticism.*
5. **elide** *v* (*linguistics*) to leave out the sound of part of a word when pronouncing it: *The 't' in 'often' may be elided.*
6. **grandeur** *n* [U] the quality of being great, magnificent and impressive in appearance or status: *the beauty and grandeur of the Swiss Alps ◦ She clearly suffers from delusions of grandeur (ie thinks she is more important than she really is).*
7. **lingua franca** *n* (*pl lingua francas*) (*usu sing*) a language used for communicating between people of an area in which several languages are spoken: *English is becoming the lingua franca of the world.*
8. **ominous** *adj* suggesting that sth bad is going to happen; threatening: *an ominous silence ◦ Those black clouds are/look a bit ominous.*
9. **repulsive** *adj* causing a feeling of strong dislike or disgust: *a repulsive sight / smell / person ◦ Picking your nose is a repulsive habit.*
10. **restive** *adj* unable to be still or quiet; difficult to control, esp because one is not satisfied with sth: *Another hour passed and the crowd grew/became increasingly restive.*

Appendix II: List of test entries (definition only)

1. **aftermath** *n* (*usu sing*) the circumstances that follow and are a consequence of an event, esp of an unpleasant one.
2. **aisle** *n* a passage between rows of seats in a church, theatre, railway carriage, etc or between rows of shelves in a shop.
3. **bona fide** *adj* [esp attrib], *adv* genuine(ly); without deception; legal(ly).
4. **deflect** *v* change or make sth change direction, esp after hitting sth.
5. **elide** *v* (*linguistics*) to leave out the sound of part of a word when pronouncing it.
6. **grandeur** *n* [U] the quality of being great, magnificent and impressive in appearance or status.
7. **lingua franca** *n* (*pl lingua francas*) (*usu sing*) a language used for communicating between people of an area in which several languages are spoken.
8. **ominous** *adj* suggesting that sth bad is going to happen; threatening.
9. **repulsive** *adj* causing a feeling of strong dislike or disgust.
10. **restive** *adj* unable to be still or quiet; difficult to control, esp because one is not satisfied with sth.

The Dilemma of Grammatical Data in Travel Dictionaries

Birger Andersen (*ba@asb.dk*) and Patrick Leroyer (*pl@asb.dk*),
*Centre for Lexicography, Aarhus School of Business,
University of Aarhus, Aarhus, Denmark*

Abstract: Travel dictionaries are edited with a view to helping international tourists with no or very little previous knowledge of the foreign language. They aim to solve communicative problems in a number of specific situations in which tourists wish to engage in conversation. However, they also have to deal with the absence of the necessary listening and speaking skills of their users. The lexicographic information is normally twofold, as it includes a selection of relevant words and sentences as well as a selection of the grammatical rules that govern the integration of these words and sentences into tourist conversation — from words to utterances so to speak. In this article, we will focus on the selection and presentation of the grammatical data in order to investigate how they contribute to the fulfilment of their primary functions. Our thesis is that the grammatical data have to deal with a dilemma: helping tourists understand and be understood during a conversation while also helping them acquire some basic knowledge of the grammatical system of the foreign language.

Keywords: TRAVEL DICTIONARIES, BILINGUAL DICTIONARIES, LEARNERS' DICTIONARIES, ORAL TEXT RECEPTION, ORAL TEXT PRODUCTION, GRAMMATICAL DATA, SYNCHRONOUS CONSULTATION, ASYNCHRONOUS CONSULTATION

Opsomming: Die dilemma van taalgegewens in reiswoordeboeke. Reiswoordeboeke word opgestel met die doel om internasionale toeriste met geen of baie min kennis van 'n vreemde taal te help. Hulle beoog om kommunikasieprobleme op te los in 'n aantal spesifieke situasies waarbinne toeriste graag in gesprek wil tree. Hulle moet egter ook die afwesigheid van die nodige luister- en spreekvaardighede van hul gebruikers hanteer. Die leksikografiese inligting is gewoonlik tweeledig omdat dit 'n keuse van toepaslike woorde en sinne insluit sowel as 'n keuse van die taalreëls wat die integrasie van hierdie woorde en sinne in toeristegesprekke bepaal — van woorde tot uitinge as 't ware. In hierdie artikel sal ons fokus op die keuse en aanbieding van taalgegewens ten einde na te gaan hoe hulle tot die verwesenliking van hul primêre funksies bydra. Ons beskouing is dat die taalgegewens 'n dilemma moet hanteer: om toeriste te help om te verstaan en verstaan te word tydens 'n gesprek, en om hulle terselfdertyd te help om ook 'n basiese kennis van die taalsisteem van die vreemde taal op te doen.

Sleutelwoorde: REISWOORDEBOEKE, TWEETALIGE WOORDEBOEKE, AANLEERDERSWOORDEBOEKE, MONDELINGE TEKSONTVANGS, MONDELINGE TEKSPRODUKSIE, TAALGEGEWENS, SINCHRONIESE RAADPLEGING, ASINCHRONIESE RAADPLEGING

1. Inheriting the grammatical characteristics of bilingual dictionaries

Functionally speaking, lexicographic tools for tourists, including the so-called travel dictionaries and phrasebooks,[†] share a large number of lexicographic characteristics with the related 'ordinary' bilingual dictionaries. They are conceived and designed to help their intended L1 native users encountering communicative problems (mostly in conversations) by providing L1–L2, and to a certain degree L2–L1, foreign language assistance in the form of equivalents to single words and multi-word expressions, and to a range of prefabricated sentences to be used in conversation in conventional tourist situations. Our study of a number of these dictionaries reveals that in almost all cases the lexicographic information also includes the distribution of a substantial amount of grammatical data which can be accessed in the situational articles and in the bidirectional, bilingual wordlists, and mostly in a specific grammar section in L1 — a so-called mini-grammar. Information is normally provided from word to sentence level — phonetics, morphology, syntax, semantics as well as style and pragmatics, and even at the cultural level.

Our first hypothesis is that in travel dictionaries, the authors' decision to include a substantial amount of grammatical data at all levels is motivated by the authors' wish to exploit and even surpass the lexicographic bilingual legacy, as the value of including grammatical data in bilingual dictionaries has been broadly acknowledged for the past thirty years. It seems therefore reasonable to assume that travel dictionaries have merely inherited the characteristics of this lexicographic tradition from a bilingual perspective. Many metalexigraphic contributions have dealt with the nature and the distribution of the grammatical data in the articles or in a certain type of dictionary grammar proper (Andersen 2007, Mugdan 1989, Mogensen 2005, Tarp 2006). The function of the grammatical data has also been discussed, and there seems to be a general agreement on the superiority of a production grammar (Andersen 2007: 133, Tarp 2006: 281): helping the L1 user to produce texts in the L2. Although it is not explicitly stated, all metalexigraphic contributions deal with written text production, and nothing is being said about oral text production in L2 and oral text reception of L2, which obviously are the two primary functions of lexicographic tools designed for tourists.

2. Inheriting the grammatical characteristics of learners' dictionaries

What truly distinguishes travel dictionaries from bilingual dictionaries is the widely accepted metalexigraphic identification of their intended users and user situations:

- users with no or very little previous knowledge of the foreign language, and
- users who need actual communicative assistance in relevant tourist situations.

While being completely in line with the above stated description of the user group profile, Abend (1991: 2903) should be credited for insisting on the dictionary as a lexicographic tool:

Das Reisewörterbuch (= RWB) ist als lexikographisches *Hilfsmittel* (our italics) für Reisende gedacht, die nur geringe oder keine Kenntnisse der Sprache ihres Urlaubslandes besitzen.

The travel dictionary as a lexicographic tool is aimed at travellers who have only very little or no knowledge of the language of their country of destination.

We could not agree more on the purpose of the travel dictionary as a tool, but find it very difficult to accept the definition of the user profile. The problem is obviously how travel dictionaries can possibly fulfil their communicative function and really help their intended users in the foreseen situations. As argued in Leroyer (2007, 2008), it is fairly obvious that tourist dictionaries are incapable of helping tourists who have no or little previous knowledge of the foreign language to communicate orally in it. That also goes for the grammatical assistance included in the lexicographic tool. To engage in a conversation in a foreign language presupposes some previous degree of foreign language knowledge and ability, including both listening competences and listening skills, and no matter how, the grammatical assistance will not help without these competences and skills. In other words, the real users of travel dictionaries should truly be identified and described as active learners of the foreign language at a beginner level, bound to encounter basic conversational problems in tourist relevant communication situations. Abend (1991: 2904) rightly claims that travel dictionaries are designed to be punctually consulted during travel:

Das RWB dient als Nachschlagewerk auf Reisen, um aktuelle Kommunikationsprobleme durch punktuelle Konsultation zu beheben.

The travel dictionary serves as a reference work on travel in order to solve current communication problems through punctual consultation.

but immediately admits that other user situations are perfectly conceivable:

RWB werden auch für die Reisevorbereitung [...] verwendet.

Travel dictionaries are also used for travel preparation.

From this consultation perspective, our second hypothesis is that the authors' decision to include substantial grammatical data in travel dictionaries must be seen as an attempt to compensate for a strongly suspected deficiency with respect to communicative assistance. This is done by giving access to the consultation of basic, pedagogically structured information on the language system and on its use: 'How the language works'. Travel dictionaries thus become hybrids: bilingual and learners' dictionaries at the same time. From a functional perspective, they are always consulted in temporally determined scenarios, in which communication and knowledge-oriented situations alternate (see also

Tarp 2006: 243-283 for a comprehensive overview of the lexicographic functions with regard to distribution of, and access to grammatical data in learners' dictionaries):

- (a) a knowledge-oriented **prospective** consultation taking place prior to travelling, in which travel dictionaries are systematically used as pedagogical tools in connection with a basic foreign language learning process,
- (b) a communication-oriented **introspective** consultation taking place while travelling, in which travel dictionaries are punctually used as practical tools in order to obtain immediate assistance with respect to current communication problems, and
- (c) a knowledge-oriented **retrospective** consultation in which travel dictionaries are systematically used to check on and eventually correct acquired knowledge, or to acquire new knowledge insofar as past encountered communication situations have revealed insufficient foreign language competence and skills (and have thus revealed insufficient dictionary assistance), or from the perspective of preparing a new travel to the same destination.

3. The consultation and the functions of grammatical data

As argued above, grammatical assistance can prove to be useful in a number of user situations in which the user punctually consults the dictionary to solve an actual communicative problem, or systematically in order to acquire the necessary knowledge on the language system and its usage as part of the language learning process. Below, we list the main interrelated lexicographic functions and consultation scenarios, in which the intended tourist user is a beginner, but has some basic previous knowledge of the foreign language, or is attending a language course:

Cognitive function — prospective consultation and foreseen general problems

- (a) asynchronous assistance — learning how to communicate in specific tourist situations as a result of foreseen general problems because of lack of foreign language competences and skills

Communicative functions — introspective consultation and actualised specific problems

- (b) synchronous assistance in connection with written text reception as a result of actual, specific problems in specific tourist situations
- (c) synchronous assistance in connection with oral text production as a result of current, specific problems in specific tourist situations

- (d) synchronous assistance in connection with oral text reception as a result of current, specific problems in specific tourist situations

Cognitive functions — retrospective consultation and experienced specific and general problems

- (e) asynchronous assistance to correct communication problems as a result of experienced specific communication problems encountered in (b)–(d)
- (f) asynchronous assistance to pursue the ongoing learning process as a result of experienced general communication problems encountered in (b)–(d), and prompted by (a)

We are now in a position to investigate the selection and presentation of grammatical data in travel dictionaries and compare them with the functions identified above in their temporally determined consultation scenarios. But first, we have to briefly describe the different levels of user competence and their associated grammatical data.

4. Eight levels of grammatical user competence from a normative perspective

Taking grammar in its broader sense, we can establish eight different linguistic competences which — if acquired and possessed by a non-native speaker — will bring him/her very close to the linguistic competences of the native speaker.

- (1) **Phonetic competence**
The speaker is capable of pronouncing words and phrases so as to make him-/herself understood. The speaker is also capable of understanding speech.
- (2) **Orthographic competence**
The speaker (writer) is capable of spelling words in correspondence with the orthographic conventions of the language.
- (3) **Formal (or morphological) competence**
The speaker is capable of correctly inflecting nouns for number, verbs for tense, adjectives for comparative and superlative, etc. etc. Formal competence also involves the following: choosing correctly (in English) between the *-ing* form and the infinitive after specific verbs, choosing the correct subject–verb agreement, choosing the correct relative pronoun, etc., etc.
- (4) **Structural (or syntactic) competence**
The speaker is capable of placing the correct number of clause and phrase constituents in the correct order in the clause and phrase. It

involves, for example, correct constituent order in inversion, correct order of premodifying adjective phrases in the noun phrase, correct order of modal and other auxiliaries in the verb phrase, etc. etc.

- (5) Semantic (or terminological) competence
The speaker is capable of choosing the correct/appropriate word in any given speech situation.
- (6) Stylistic competence
The speaker is capable of selecting lexical and grammatical elements which are appropriate for any given speech situation. The selected lexical and grammatical elements match the levels of formality and politeness required by the speech situation.
- (7) Pragmatic competence
The speaker is capable of selecting the appropriate speech act and phrasing it correctly to accomplish his/her pragmatic goals in any given speech situation.
- (8) Cultural competence
The speaker is capable of selecting the appropriate speech act and phrasing it correctly while using the appropriate body language conventions of the destination country in a particular speech situation.

There is some overlap between (6) and (7) since levels of politeness and formality are interrelated with choices between direct and indirect speech acts. There is also some overlap between (4) and (7), since a change in constituent order in a given speech act may create a different speech act (for example from statement to question). Finally, there is some overlap between (6), (7) and (8), since cultural conventions in all cases influence the speakers' choice of appropriate cultural phrases.

Linguistically speaking, these normative levels are all represented in tourist dictionaries (1)–(4) ((1)–(7) and (8) in the case of 'exotic' languages) as categories of information. Lexicographically speaking, the question, however, is the capacity of the information given by these lexicographic data to bring the necessary assistance to the communicative needs and problems of the tourist users. In the following, we will study the selection and presentation of the grammatical information.

5. The shortcomings of grammatical data from a functional perspective

5.1 Helping with the sounds

Phonetic information is given in all the dictionaries investigated. The information is normally accessible both in the articles and in the pronunciation section

of the mini-grammars. In the Danish dictionaries, this information firstly takes the form of pronunciation tables using the International Phonetic Alphabet (IPA). In the *Gyldendal* dictionaries (GSEP 2005, GSFP 2005), there are also transcriptions (using the IPA) of all the phrases and all the words in the dictionaries. The *Politiken* dictionaries (PER 2005, PFR 2005, FDDFMO 2002) only have occasional transcriptions of words in the dictionaries, presumably those the writers of the dictionaries feel present special pronunciation difficulties.

We must assume that the use of the IPA will be a problem for those tourists who have had no formal language training. Many of the phonetic symbols of the IPA do not resemble letters in any Romance or Germanic language at all. It would therefore require a very thorough study if the tourist were to make sense of these pronunciation tables and the transcriptions.

In the English dictionaries, there are also pronunciation tables and transcriptions of phrases and words. However, the transcriptions used here are 'homemade' and different from one dictionary to the next. One dictionary has the following to say about the transcription system used:

In this dictionary, the German has been written in a system of imitated pronunciation so that it can be read as though it were English. (RGG 2006: 241)

The problems with these transcription systems can be illustrated by a few examples. The French term 'Saucisse de Strasbourg' (beef sausage) is transcribed in the following ways:

RGF 2006: [sossees duh strazboorg]
LPF 2003: [so•sees der straz•boor]

And the German term 'Bauernfrühstück' (bacon and potato omelet) is transcribed as follows:

RGG 2006: [bownern-frooshtook]
LPG 2003: [bow•ern•frü•stück]

Since none of these transcriptions correspond to any orthographic pattern in English (see Wen-Chao Li 2007 for a critical survey of the many problems linked to the choice of adequate transliteration systems), it is indeed doubtful whether they will be of any assistance to tourists with respect to producing a correct or even comprehensible pronunciation of the terms. The problem is compounded when tourists have to struggle with transcriptions of longer phrases such as the following from LPF (2007: 136):

[es•ker la pray•pa•ra•syon va pron•drer bo•koo der tom] 'Est-ce que la préparation va prendre beaucoup de temps'

So far, we have looked at pronunciation and transcription only from the productive perspective. A number of the dictionaries also attempt to provide

assistance with pronunciation from a receptive angle, sometimes under headings such as 'Listen for' (*Lonely Planet*), sometimes as part of whole conversational exchanges as in the *Rough Guide* dictionaries, as for example the following from RGF (2006: 23):

hello
 ▷ bonjour
 [bONjoor]
 bonjour, je m'appelle Marie ◁
 [bONjoor juh mapel maree]
 hello, my name's Marie
 Graham, from England, Thirsk
 ▷ Graham, de Thirsk en Angleterre
 [graham, duh thirsk ON nONgluhtair]
 je ne connais pas, où est-ce? ◁
 [juh nuh konay pa, oo ess]
 don't know that, where is it?
 not far from York, in the North; and you?
 ▷ pas très loin de York, dans le nord; et vous?
 [pas tray lwan duh york dON luh nor; ay voo]
 je suis de Pau; vous êtes ici tout seul? ◁
 [juh swee duh po; voo zet-see too surl]
 I'm from Pau; here by yourself?

Apart from the fact that the tourist still has to struggle with the 'homemade' transcription system, the conversational exchange is so idiosyncratic that neither the productive nor the receptive elements have any general value whatsoever.

Either too complex, misleading or idiosyncratic, the phonological data accessible in the articles fail to help tourists solve the actual communicative problems attributable to the production and recognition of the sounds of the foreign language (functions (b) and (c)) in the case of punctual consultation situations. In the case of systematic consultation situations (functions (a), (e) and (f)), the data accessible in the pronunciation section of the mini-grammars are far too condensed to fulfil their cognitive function and help the tourist user acquire a satisfactory level of basic phonological competence. As poetically stated by the LPEA (2002: 11), descriptive, representative knowledge of the sounds of Amharic Ethiopian is simply not enough; it must be complemented by perceptive knowledge acquired through practice in order to develop adequate speaking and listening skills:

Tune into the sounds of Ethiopia's streets for tips on pronunciation: a mini-bus worker trilling his *r* as he calls the destination, **a.rat ki.lo!**: a vendor clicking her *k*'s as she advertises her spicy snacks, **ko.lo, ko.lo!** [...] with a bit of practice, you'll be told you're **gob.beuz**, 'brilliant', by native Amharic speakers.

5.2 Helping to spell correctly

Orthographic information is irrelevant in our context, since we are concerned only with oral production and reception. In actual fact, none of the dictionaries investigated give any explicit orthographic information, although it could be relevant for function (b), written text reception, particularly in the case of languages using other script systems, such as syllabary in LPEA 2002.

5.3 Helping to inflect correctly

Explicit morphological information is given in the so-called 'mini-grammars' or 'How the language works' sections of the dictionaries. We find there the essential inflectional rules with respect to the plural ending of nouns, the comparative and superlative endings of adjectives (where relevant), gender endings of nouns and adjectives (where relevant), information about subject-verb agreement, etc. etc. In this respect, it seems as if the dictionaries aim to equip the tourist with the means to express him-/herself in a language characterised by formal grammatical correctness.

However, some dictionaries go far beyond the needs of the average tourist in this respect. US 2004, for example, provides the tourist with a very thorough and intricate overview of the system of verb inflections in Spanish. The overview first divides the Spanish verbs into three inflectional groups and then gives the inflectional forms for infinitive, present and past participle, present indicative, present continuous, past simple, future, present perfect, past perfect and imperfect past. For all the finite categories we get the forms for the whole person range. This is followed by the whole range of forms for the auxiliary verbs *ser*, *estar* and *tener* (but not *haber*), and finally six frequent irregular verbs, although here the information given is for some reason restricted only to infinitive, present indicative, past simple, imperfect past and future, and for the finite categories only to 1st person singular. Other dictionaries are more restricted and focused in this respect, although the inflectional information given is still beyond what the average tourist needs.

Much in line with conventional bilingual lexicography in French (cf. Hovmark 2005), the FDDFMO travel dictionary brings information on the regular inflectional patterns of French verbs in an outside matter:

1. the *-er* group inflectional pattern
2. the *-ir* group inflectional pattern
3. the *-re* group inflectional pattern
4. the inflectional pattern of the auxiliary verb *être* (be)
5. the inflectional pattern of the auxiliary verb *avoir* (have)

as well as an alphabetical list of the following more or less irregular verbs (6 to 108):

6. absoudre (absolve)	41. dire (say)	76. peindre (paint)
7. accueillir (welcome)	42. dissoudre (dissolve)	77. percevoir (perceive)
8. acheter (buy)	43. dormir (sleep)	78. peser (weigh)
9. acquérir (acquire)	44. écrire (write)	79. placer (place)
10. aller (go)	45. envoyer (send)	80. plaindre (pity)
11. apercevoir (catch sight of)	46. essuyer (dry)	81. plaie (pleasure)
12. appeler (call)	47. exclure (exclude)	82. pleuvoir (rain)
13. apprécier (appreciate)	48. faillir (fail)	83. pourvoir (provide)
14. assaillir (assault)	49. faire (do, make)	84. pouvoir (can)
15. s'asseoir (sit)	50. falloir (have to)	85. prendre (take)
16. assiéger (besiege)	51. foutre (do)	86. recevoir (receive)
17. battre (hit)	52. fuir (escape)	87. se repentir (repent)
18. boire (drink)	53. haïr (hate)	88. résoudre (solve)
19. bouillir (boil)	54. inclure (include)	89. rire (laugh)
20. broyer (grind)	55. interdire (forbid)	90. rompre (break)
21. céder (give up)	56. jeter (throw)	91. savoir (know)
22. clore (close)	57. joindre (join)	92. sentir (feel)
23. concevoir (conceive)	58. lire (read)	93. servir (serve)
24. conclure (conclude)	59. luire (shine)	94. sortir (go out)
25. confire (candy)	60. manger (eat)	95. souffrir (suffer)
26. connaître (know)	61. maudire (curse)	96. suffire (be enough)
27. contraindre (force)	62. se méfier (mistrust)	97. suivre (follow)
28. convaincre (convince)	63. mentir (lie)	98. taire (be silent)
29. coudre (sew)	64. mettre (put)	99. tenir (hold)
30. courir (run)	65. modeller (model)	100. traire (milk)
31. couvrir (cover)	66. moulin (grind)	101. tressaillir (shudder)
32. craindre (fear)	67. mourir (die)	102. vaincre (defeat)
33. créer (create)	68. mouvoir (move)	103. valoir (be worth)
34. croire (believe)	69. naître (be born)	104. venir (come)
35. croître (grow)	70. nuire (harm)	105. vêtir (dress)
36. cueillir (pick)	71. offrir (offer)	106. vivre (live)
37. cuire (cook)	72. ouvrir (open)	107. voir (see)
38. décevoir (disappoint)	73. paraître (seem)	108. vouloir (want)
39. défaillir (faint)	74. partir (leave)	
40. devoir (must)	75. payer (pay)	

The list contains information on the irregular forms used in the following tenses: present, perfect, imperfect, future, conditional and subjunctive. The full conjugation pattern is only given for the present tense, whereas nothing is explained about the function of the other forms and tenses (expressing differences of time, aspect and modality). The real problem however is not the fact that (complete) grammatical information is lacking but that the grammatical data type 'irregular verbs and their inflections' is functionally speaking inconsequential and irrelevant to the tourist user. The verbs are for many different reasons useless in conventional tourist user situations, either because their usage is archaic (22: *clorre*), extremely vulgar (51: *foutre*), very rare (35: *croître*, 68: *mouvoir*, 87: *se repentir*, 101: *tressaillir*), highly specialised (6: *absoudre* in the field

of religion, 16: *assiéger* in the field of military tactics, and 100: *traire* in the field of cattle breeding), or simply because they are only inflected in one single form (66: *moudre* → *moulu*, when speaking about coffee). As already pointed out in the case of the Spanish verb examples, the information given there goes far beyond the functional needs of the average tourist with regard to formal grammatical correctness and can merely be interpreted as a case of lexicographic conservatism resulting in lexicographic cloning.

To sum up, it appears, functionally speaking, that the representation of inflectional data far exceeds the average tourist needs in punctual (functions (c) and (d)) as well as in systematic consultation situations ((a), (e) and (f)). The inflectional data also take up considerable space, consequently severely reducing the amount of grammatical data needed.

5.4 Helping to construct sentences

In general, the mini-grammars of the dictionaries do not give much information about syntactic matters. There are however occasional remarks on matters such as 'word order', for example:

I'm travelling to Berlin

Ich fahre nach Berlin
(lit: I travel to Berlin)

Tomorrow I'm travelling to Berlin

Morgen fahre ich nach Berlin
(lit: tomorrow travel I to Berlin)

In a straightforward German statement, the verb is the second element (note that this does not necessarily mean it's the second word as the first element can contain more than one word). The verb would usually follow the subject, i.e., the concept or thing you're talking about. So if you start the sentence with a word like 'tomorrow', you have to reverse the order of the subject and verb to keep the verb in second spot, as illustrated above. (LPG 2003: 24)

Or:

I'm looking for a comfortable hotel

Je cherche un hôtel confortable
(lit: I look-for a hotel comfortable)

As a rule, adjectives come after the noun in French. There are exceptions, however, two useful ones to know being 'big' (*grand*) and 'small' (*petit*) which come before the noun. (LPF 2003: 16)

Otherwise, in this respect, the tourist is left to deductive reasoning from the phrases. However, in most cases it is not essential for successful communication that every sentence contains the correct number of clause constituents,

whereas correct constituent order is in many cases essential in, for example, distinguishing between the statement of a declarative sentence and the question of an interrogative sentence. This aspect will be further discussed under 5.7 below.

5.5 Helping to pick up the right word or the right phrase

The dictionaries provide explicit semantic information for productive purposes in the form of general $L_1 \rightarrow L_2$ dictionaries and a number of thematic $L_1 \rightarrow L_2$ dictionaries, such as for example a list of useful words if you stay at a youth hostel (PER 2005: 69) or the words you need to describe problems with your car (US 2004: 57-58).

They also provide explicit semantic information for receptive purposes in the form of general $L_2 \rightarrow L_1$ dictionaries and again a number of thematic $L_2 \rightarrow L_1$ dictionaries such as a list of terms you may find on menus in restaurants (GSEP 2005: 112-118), or words you may hear from the mechanic in case your car breaks down (US 2004: 59-60), under the assumption, that is, that you have studied the transcriptions of the words and phrases so as to be capable of recognizing for example the following:

[leemptyah-pahrah**breesahs**]

which is 'Spanish' for *windshield wiper*.

The approximate number of entries in the general dictionaries and in the thematic dictionaries for tourists with France as their destination appears from the table below:

	English/Danish→French	French→English/Danish
<i>Lonely Planet</i>	2 600	2 500
<i>Rough Guide</i>	4 400	3 400
<i>Gyldendal</i>	2 900	2 800
<i>Politiken</i>	1 700	1 000

It should be observed, however, that whereas the entries in the thematic dictionaries are on the whole included in the general dictionaries *Lonely Planet*, *Rough Guide* and *Gyldendal*, this is not the case with *Politiken*. The total number of entries in *Politiken* is therefore probably the same as in *Lonely Planet* and *Gyldendal*.

In general, the selection of entries seems to be appropriate, for the words and phrases selected will — to a large extent — cover the semantic-communicative needs of the average tourist, thus more or less fulfilling functions (c) and (d). One may wonder, however, how words such as *activist*, *antinuclear* and *deforestation* have found their way into the general $L_1 \rightarrow L_2$ *Lonely Planet* dictionary. It can hardly be imagined how a tourist with a very limited knowledge of

French would engage in communicative situations that would require the use of such words.

5.6 Helping not to sound rude

Stylistic information helps tourists select words and phrases that fit in with the given communicative situation in terms of levels of politeness and formality. Such information is actually scarce in the dictionaries investigated, but there are occasional remarks such as the following from the section on pronouns in the chapter 'How the language works' in RGF (2006: 245):

YOU: **tu** is used when speaking to someone who is a friend, or to someone of your own general age group with whom you want to establish a friendly atmosphere. **Vous** is used when speaking to several friends (i.e. it is the plural of **tu**) or when speaking to someone you don't know. In the vast majority of cases, as a foreigner in France you will use the **vous** form. Certainly, if you are in any doubt as to which form to use, choose the **vous** form.

The following from the chapter 'How the language works' in RGG 2006, the section on modal verbs in the mini-grammar in PER 2005, and the section 'Language Difficulties' in LPG 2003 fall into the same category:

In the English–German section, when two forms of the verb are given in phrases such as 'can you ...? **kannst du / können Sie ...?**' the first is the familiar and the second the polite form ... (RGG 2006: 241)

Man bruger ofte datidsformerne når man henvender sig til nogen på en høflig måde:

The past tense forms are often used when approaching somebody in a polite way:

Could you tell me when the bus leaves?

Kan De sige mig hvornår bussen går?

Would you like a cup of tea?

Vil du have en kop te? (PER 2005: 180)

There are two forms of the second person singular pronoun 'you'. Use the polite form *Sie* with anyone you don't know well. You should only use the informal form *du* with people you know very well. All the phrases in this chapter use *Sie* — use your intuition to work out when to use formal and informal forms. (LPG 2003: 25)

Occasionally, the dictionaries will provide information about the level of formality of the entry. The entry 'appointment' in the general $L_2 \rightarrow L_1$ dictionary of GSEP 2005 is thus given the label 'formal'.

It is somewhat surprising that dictionaries, which devote whole chapters

to an explicit treatment of grammatical correctness, do very little (explicitly) to equip the tourist with the linguistic means with which to select words and phrases that fit the given communicative situation in terms of politeness and formality. One explanation for this could be that the writers of the dictionaries feel that they have dealt sufficiently implicitly with this aspect in their selection of phrases in the situational-contextual chapters. It can therefore be concluded that the grammatical data assigned to information on formality and politeness fail to fulfil their intended functions at the stylistic level: helping tourists not to sound rude (functions (c) and (d)).

5.7 Helping to do things with words

Information about pragmatics comes in various forms in the dictionaries. The main justification for the dictionary is in fact to equip the tourist with a number of speech acts which, if used appropriately and phrased correctly, will accomplish his/her communicative goals in any given speech situation. The following examples of speech acts all come from LPF 2003.

One typical speech act is 'requesting information':

What level is that slope?

Quelle est la difficulté de cette piste?

Another is 'requesting a (physical) thing':

Same again please.

La même chose, s'il vous plaît.

It could be an 'offer':

I'll buy you a drink.

Je vous offre un verre.

Or a 'compliment':

You're a fantastic dancer.

Tu dances vraiment bien.

However, we have some difficulty in classifying the following phrase found under the section 'Romance' in LPG (2003: 113). It is particularly unclear what other professions could be inserted into the parentheses with the same perlocutionary effect, whatever it is:

Before this goes any further, I must be upfront. I'm an (accountant).

Bevor wir uns näher kennen lernen, muss ich etwas klarstellen. Ich bin (Buchhalter/Buchhalterin).

All of these of course give implicit information about the performance of speech acts in the foreign language, but very often the dictionaries will also provide explicit information about speech acts and how to perform them.

The following from PFR (2005: 204) illustrates explicit instruction in how to phrase a specific speech act:

SPØRGSMÅL
QUESTIONS

Et spørgsmål som fx *Har De et værelse?* kan formuleres på tre måder:
A question such as *Har De et værelse?* can be formulated in three ways:

1. **Avez-vous** une chambre? dvs. ved omvendt ordstilling ligesom på dansk.
i.e. by inversion like in Danish.
2. **Vous avez** une chambre? dvs. ved blot at hæve tonen sidst i sætningen (ligesom man kan gøre i nogle spørgsmål på dansk).
i.e. by raising intonation at the end of the sentence (in the same way as it is possible in some questions in Danish).
3. **Est-ce que vous avez** une chambre? dvs. ved at sætte spørgefrasen **est-ce que**, som udtales [æske], foran.
i.e. by putting the question phrase **est-ce que**, which is pronounced [æske] in front.

Similar explicit instruction is found in LPG (2003: 23-24):

yes / no questions

The hotel is on Potsdamer Square?

Das Hotel ist am Potsdamerplatz?
(lit: the hotel is on-the Potsdamer-square)

There are three main ways of forming yes/no questions. The one which requires least effort is to make a statement, but to say it like a question, rising in intonation towards the end of the sentence, as illustrated above.

The hotel is on Potsdamer Square, isn't it?

Das Hotel ist am Potsdamerplatz, nicht wahr?
(lit: the hotel is on-the Potsdamer-square not true)

The second example shows that you can also add *nicht wahr* (not true) to the end of a statement. This is much easier than in English, where you'd use a variety of question tags, such as 'isn't it?', 'aren't you?' and 'doesn't it?'

Is the hotel on Potsdamer Square?

Ist das Hotel am Potsdamerplatz?
(lit: is the hotel on-the Potsdamer-square)

Finally, the third example shows you can turn a statement into a question by reversing the order of the subject and the verb.

The literal illustration of word order in the syntactic construction of L2 sentences is sometimes deliberately elucidated in the articles of the phrase book, as in LMP 2004. The examples below are taken from a conversational section containing sentences that might be useful in a bargaining situation (from the

very first line 'how much is this sculpture' to the last line 'here is your money, don't you have any change?'):

Lafo kosa izany ka!

lafou koussa zani ka

cher quand-même cela hein (= expensive however this eh)

Que c'est cher!

Tsy lafo iazany sady tsara tarehy koa anie io e!

tssi lafou zani ssadi tssara taré kou ani-é i-ou é

ne-pas cher cela est beau aspect aussi voyez-vous ceci (= not expensive this is beautiful aspect also see-you this)

Ce n'est pas cher et elle est belle n'est-ce pas?

As already argued above, the data included in the 'example of a complete bargaining conversational exchange' are so idiosyncratic that they do not fulfil any productive or receptive function whatsoever (functions (c) and (d)). The lack of functional usability also applies to the data type 'L1 translation of literary L2 word order' in association with the 'natural L2-L1 translation' data type. This type of data selection and presentation merely reproduces the conventions of a linguistic style sheet tradition in journals of linguistics, in which it is necessary that the reader gains direct access to knowledge on the syntactic constraints of the so-called exotic languages in terms of word order and grammatical categories. As the French-Malagasy is intended for the tourist (cf. the foreword p. 1) and not for the linguist, and as the grammatical rules given for word order in the mini-grammar of the book (pp. 13-14) are clearly insufficient, the only lexicographic function we can think of is a knowledge-oriented function, namely contrastive knowledge of word order in Malagasy, which clearly fails to fall in under cognitive functions (e) and (f). Furthermore, the words used in the sentence examples of the situational part are not even lemmatised in the bilingual word list, the phrase book consequently failing here to assist with any possible reception problems.

An important aspect of pragmatic competence is knowledge of how to engage in 'phatic communion' and knowledge of the essential politeness phrases. These aspects are covered extensively in all the dictionaries investigated. GSE 2005 for example has a whole chapter (17 pages in length) with examples of how to greet people, politeness phrases, phrases expressing opinions and emotions, compliments, small talk, etc.

An essential problem with all this is, however, to what extent this plethora of phrases is essential to the **basic communicative needs** of the tourist. It is not essential for the tourist who walks into a, say, French pharmacy to buy painkillers to be conversant with a whole range of different ways of expressing the speech act 'requesting a thing'. The conventions of the social and communicative situation in which he/she finds him-/herself will in fact define almost all aspects of the situation, so that all he/she needs to do is to utter the French word for *painkillers*. The conventions of the situation will see to the rest.

5.8 Helping to speak and behave correctly

The following data from the 'savoir vivre' section of LMP 2004 illustrate explicit instructions in how to phrase a specific speech act and how to use the appropriate body language in the context of greetings and politeness — here to show respect to the elderly:

S'il vous arrive de passer devant une personne âgée, vous devrez dire **azafady tompoko**, *s'il vous plaît monsieur/madame*, tout en vous courbant un peu.

If you happen to pass before an elderly person, you must say **azafady tompoko**, *please sir/madam*, while slightly bowing.

The problem, however, is that these important cultural instructions and phrases are not integrated in any way in the distribution structure of the dictionary (situational sections and/or bidirectional wordlists) and are thus not easily accessible to the user (see Gouws 2007 for functional propositions with regard to the microstructural integration of cultural data types).

6. Conclusions — mixing the characteristics of bilingual and learners' dictionaries

Our investigation of travel dictionaries has revealed the functional dilemma of grammatical data which jeopardises the efficiency of lexicographic tools designed for tourists at a beginner level. As symptomatically stated by the RGF 2006 in its foreword — 'The RGF dictionary is a highly practical introduction to contemporary language' — travel dictionaries seem to be perfectly aware of their double identity, simultaneously bilingual dictionaries and learners' dictionaries. They intend to be used as practical tools, and yet do not admit frankly that their usability is subject to their potential users' participation in active language learning processes. In other words, one might say that travel dictionaries somehow do not attempt to free themselves from their pedagogical constraints.

The functional dilemma then lies in the fact that communicatively oriented grammatical assistance is determined by punctual consultation scenarios, whereas cognitively oriented grammatical assistance is determined by systematic consultation scenarios triggered by the shortcomings of the communicative assistance; accordingly it lies in the fact that the two consultation scenarios, while being interdependent, are temporally determined.

Our study has also revealed the overall predominance of the linguistic approach, with a disproportionate focus on correctness. It means that no clear choice is made with regard to the nature of the lexicographic assistance provided by the grammatical data: the two sets of orientations remain separated instead of being integrated in the overall data distribution structure. Travel dictionaries are caught between the bilingual and the pedagogical legacy of

lexicography for users of a foreign language, between cognition and communication; they want to do both, but fail to deal with the asynchronicity of their intertwined orientations.

The resolution of the dilemma of grammatical data lies in mixing the lexicographic characteristics and conceiving bilingualised learners' dictionaries which are less ambitious with regard to linguistic correctness, but far more ambitious with regard to the integration of pedagogical grammatical data for language learning and practising. In order to ensure flexible and easy access to the grammatical data, this development could best be achieved by the publication of online travel dictionaries linked to associated online basic language courses, whereby consultation asynchronicity would truly become a functional asset.

Note

- † In this article, we will simply use the term 'travel dictionary' and abandon the term 'phrase book', much in line with Bergenholtz and Bergenholtz (2007) who argue: 'The titles of lexicographic tools — dictionary, lexicon or encyclopaedia — are normally just names given to them by the publishers, although you may point out certain differences between certain types.' In our case, there are no differences, as both terms refer to types of lexicographic tools designed for tourists in order to provide the same kind of lexicographic assistance with the same kind of lexicographic resources.

Literature

Dictionaries

- Assimil Évasion. 2004. *Le Malgache de poche*. (LMP)
- Gyldendals Store Engelsk Parlør. 2005. Copenhagen: Gyldendalske Boghandel, Nordisk Forlag A/S. (GSEP)
- Gyldendals Store Fransk Parlør. 2005. Copenhagen: Gyldendalske Boghandel, Nordisk Forlag A/S. (GSFP)
- Lonely Planet French. 2003. Second edition. Victoria: Lonely Planet Publications. (LPF)
- Lonely Planet German. 2003. Second edition. Victoria: Lonely Planet Publications. (LPG)
- Lonely Planet Ethiopian Amharic Phrase Book. 2002. Victoria: Lonely Planet Publications (LPEA)
- Politikens Engelsk Rejseparlør. 2005. Copenhagen: Politikens Forlag. (PER)
- Politikens Fransk Rejseparlør. 2005. Copenhagen: Politikens Forlag. (PFR)
- Politikens Miniordbog Fransk. 2002. Fransk–Dansk/Dansk–Fransk Miniordbog. Copenhagen: Politikens Forlag. (FDDFMO)
- Rough Guide French. 2006. Second edition. London: Rough Guides Ltd. (RGF)
- Rough Guide German. 2006. Second edition. London: Rough Guide Ltd. (RGG)
- Ulysses Spanish for Better Travel in Spain. 2004. Ulysses Travel Guides. (US)

Theoretical contributions

- Abend, Heike.** 1991. Das Reisewörterbuch. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989–1991: 2903-2908.
- Andersen, Birger.** 2007. Dictionary Grammars. *Hermes, Journal of Language and Communication Studies* 38: 119-136.
- Bergenholtz, Henning and Inger Bergenholtz.** 2007. Kvaliteten af de leksikografiske definitioner i ordbøger, leksika og encyklopædier. *LexicoNordica* 14: 11-33.
- Gouws, Rufus.** 2007. The Selection, Presentation and Treatment of Cultural Phrases in a Multicultural Dictionary. *Lexicographica* 22: 24-36.
- 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. Third volume.* Handbücher zur Sprach- und Kommunikationswissenschaft 5.3. Berlin/New York: Walter de Gruyter.
- Hovmark, Henrik.** 2005. Tre franske miniordbøger til rejsebrug og daglig tale. *LexicoNordica* 12: 213-231.
- Leroyer, Patrick.** 2007. Maultasche og Novillada: En teori for turistordbøger. *Nordiske studier for leksikografi 9. Konference om leksikografi i Norden, Akureyri, Island, 23–24 maj 2007.* [In print.]
- Leroyer, Patrick.** 2008. Du support d'information à l'outil lexicographique: la lexicographisation du guide touristique. Bernal, Elisenda and Janet DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008:* 879-886. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Mogensen, Jens Erik.** 2005. Grammatik im Wörterbuch. Barz, Irmhild, Henning Bergenholtz and Jarmo Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch:* 189-202. Finnische Beiträge zur Germanistik 14. Frankfurt a.M.: Peter Lang.
- Mugdan, Joachim.** 1989. Grundzüge der Konzeption einer Wörterbuchgrammatik. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.). 1989–1991: 732-749.
- 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: Max Niemeyer.
- Wen-Chao Li, Chris.** 2007. Foreign Names into Native Tongues. How to Transfer Sound between Languages — Transliteration, Phonological Translation, Nativization, and Implications for Translation Theory. *Target* 19(1): 45-68.

Voorstelle vir die produksie van 'n vertaalwoordeboek vir professionele vertalers*

M.F. Crafford (*riacrafford@adept.co.za*) en Ilse Feinauer (*aef@sun.ac.za*)
*Departement Afrikaans en Nederlands, Universiteit van Stellenbosch,
Stellenbosch, Republiek van Suid-Afrika*

Opsomming: Die profiel van die potensiële gebruiker moet die riglyn wees by alle leksikografiese besluite telkens wanneer woordeboeke beplan, hersien of saamgestel word. Die formaat, struktuur en ordening van 'n vertaalwoordeboek gerig op professionele vertalers behoort gevolglik dus op die spesifieke behoeftes van hierdie bepaalde gebruikersgroep gebaseer te wees. Hierdie artikel bevat algemene riglyne vir die hantering van al twee die buitetekste en die sentrale lys van 'n vertaalwoordeboek gerig op die behoeftes van beroepsvertalers wat hoofsaaklik uit Engels (T2) in Afrikaans (T1) vertaal. 'n Moontlike formaat vir die mikrostruktuur van so 'n woordeboek (Engels > Afrikaans) word gegee en die moontlike leksikografiese bewerking van 'n aantal voorbeeldlemmas word in besonderhede bespreek. Die moontlikheid van formatiewe evaluering met behulp van die beginsel van gelyktydige terugvoer word kortliks aangeraak.

Sleutelwoorde: WOORDEBOEKPRODUKSIE, VERTAALWOORDEBOEK, WOORDEBOEKINSKRYWINGS, WOORDEBOEFUNKSIES, LEKSIKOGRAFIESE ETIKETTE, GELYKTYDIGE TERUGVOER, WOORDEBOEKKORPUS, MAKROSTRUKTUUR, MIKROSTRUKTUUR

Abstract: Suggestions for the Production of a Translation Dictionary for Professional Translators. The profile of the potential user needs to be the guiding principle in all lexicographic decisions whenever dictionaries are designed, revised or compiled. It follows, then, that the format, structure and organisation of a translation dictionary aimed at the needs of professional translators should be based on the specific needs of that distinct user group. This article contains general guidelines for both the outer texts and the central part of a translation dictionary aimed at the needs of professional translators mainly translating from English (T2) into Afrikaans (T1). A possible format for the microstructure of such a dictionary (English > Afrikaans) is given and the possible lexicographic treatment of a few sample entries is discussed in detail. The possibility of formative evaluation by means of the principle of simultaneous feedback is briefly touched on.

Keywords: DICTIONARY PRODUCTION, TRANSLATION DICTIONARY, DICTIONARY ENTRIES, DICTIONARY FUNCTIONS, LEXICOGRAPHIC LABELLING, SIMULTANEOUS FEEDBACK, DICTIONARY CORPUS, MACROSTRUCTURE, MICROSTRUCTURE

* Hierdie artikel is gebaseer op 'n M.Phil.-skripsie *Vertalers en hul bronne: Die behoefte aan 'n vertaalwoordeboek met Engels en Afrikaans as behandelde taalpaar* wat in 2005 onder leiding van prof. A.E. Feinauer aan die Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika, voltooi is.

1. Inleiding

Roberts (1990: 92) wys op die gebrek aan gepubliseerde materiaal met betrekking tot 'n metodologie van woordeboekproduksie. Sy meen dis onder meer omdat nuwe woordeboeke dikwels uit voriges ontwikkel. Hulle ontstaan sodoende op grond van veranderinge wat aan hul voorgangers gemaak word. As direkteur van die Bilingual Canadian Dictionary (BCD) Project,¹ is Roberts (1996: 16-21) wel in staat om op grond van haar eie ervaring van woordeboekmaak bepaalde stappe in die beplanningsproses te beskryf wat belangrik is wanneer 'n nuwe woordeboek beslag moet kry. Die stappe wat sy uiteensit, dui op belangrike besluite wat oor verskeie aspekte van 'n nuwe woordeboek geneem moet word. Gaan dit byvoorbeeld slegs in gedrukte vorm of ook elektronies beskikbaar gestel word? Louw (1999: 108) beklemtoon dat die sukses van woordeboeke afhang van die toeganklikheid van hul makro- en mikrostrukturele inligting. Hy toon hoedat elektroniese woordeboeke geleenthede vir die implementering van gidsstrukture skep wat nie in gedrukte weergawes moontlik is nie. In die lig hiervan is dit gewens dat 'n vertaalwoordeboek ook in 'n elektroniese weergawe verskyn — 'n besluit wat by vertalers byval behoort te vind.

Navorsing oor die vertaalproses en die werkwyse en eienskappe van professionele vertalers toon dat dié groep potensiële woordeboekgebruikers eiesoortige behoeftes het wat nie noodwendig optimaal deur tweetalige woordeboeke, soos hulle tradisioneel daar uitsien, bevredig word nie (Crafford 2005). Omdat die belangrikste uitgangspunt by die beplanning, hersiening of samestelling van woordeboeke die potensiële woordeboekgebruiker moet wees, moet professionele vertalers se behoeftes die grondslag vorm vir alle besluite oor die formaat, struktuur en ordening van 'n vertaalwoordeboek vir dié gebruikersgroep. Meer nog, professionele vertalers behoort deel te wees van die span leksikograwe wat aan so 'n vertaalwoordeboek werk.

'n Belangrike eerste stap is dat die span leksikograwe oor die algemene aard van die beoogde woordeboek duidelikheid kry. Vir die doeleindes van hierdie artikel word gekonsentreer op 'n vertaalwoordeboek wat primêr op die behoeftes van professionele vertalers uit Engels (T2) in Afrikaans (T1) werk. Die potensiële gebruikers van so 'n vertaalwoordeboek in Engels en Afrikaans is 'n gesofistikeerde gebruikersgroep wat oor 'n goeie kennis van sowel hul tweede as hul eerste taal beskik en meer inligting benodig as wat in gewone tweetalige woordeboeke gegee word. Die Engels-Afrikaans-vertaalwoordeboek behoort 'n algemene woordeboek te wees "[which] will also include a fairly high proportion of terms in a number of fields, as well as word combinations such as collocations and fixed expressions often found in special-purpose dictionaries" (Roberts 1996: 17). Beperkings ten opsigte van die omvang van die woordeboek moet in ag geneem word. Hoe meer inligting, hoe dikker die woordeboek.

In aansluiting by die verskillende kenmerke waaroor 'n vertaalwoorde-

boek ideaal gesproke moet beskik (Crafford 2005), word vervolgens aan die hantering van die voortekste-afdeling, sentrale lys en agtertekste-afdeling aandag gegee. Hedendaagse woordeboeke wat "draers van tekssoorte" is, gebruik die sogenaamde raamstruktuur om die verskillende datatipes in 'n woordeboek te akkommodeer (Gouws 2001: 81). Deeglike beplanning is nodig vir die strukturering en inhoud van sowel die sentrale lys as die buitetekste. Die buitetekste wat nie net tot beter inligtingsontsluiting vir die woordeboekgebruiker bydra nie, maar ook 'n rol in die dataverspreidingstruktuur van die woordeboek speel (Keyser 2003: 35-36), moet as deel van die omvattende leksikografiese proses aandag kry.

2. Die voortekste-afdeling

Die voortekste-afdeling van 'n vertaalwoordeboek sal uit elemente soos die volgende bestaan:

- 'n Voorwoord wat onder meer aandui waarom en vir wie die woordeboek saamgestel is, waarom bepaalde besluite geneem is en watter metodes die samestellers gevolg het.
- Verklarende aantekeninge wat verduidelik hoe gebruikers woorde en uitdrukkings kan vind waarna hulle soek, wat die redaksionele afkortings beteken, hoe verskille in spelling of betekenis tussen Britse en Amerikaanse Engels aangedui word, en hoe die inligting georden is.
- 'n Gebruikersgids wat die toegang-, artikel- en mediostruktuur van die woordeboek uiteensit en verduidelik watter struktuurmerkers gebruik word.
- 'n Afdeling waarin leksikografiese gebruike verduidelik word, soos gebruiksetikette, kruisverwysings, lettergreepverdeling en uitspraakleiding.

Die beoogde gebruiker en die funksie(s) van die woordeboek bepaal in watter taal die inligting in die buitetekste gegee word. In 'n Engels-Afrikaans-vertaalwoordeboek sal die inligting in die doeltaal wees, m.a.w. in Afrikaans.

3. Die sentrale lys

Die sentrale lys van die beoogde Engels-Afrikaans-vertaalwoordeboek sal uit die artikelreeks van A-Z saamgestel wees.² Elke woordeboekartikel bestaan uit 'n lemma as die vernaamste bewerkingseenheid en inskrywings wat deel van die bewerking daarvan uitmaak.

Belangrike besluite wat die leksikografiese span moet neem, handel oor die makrostruktuur (die versameling lemmas). Watter leksikale items moet in

'n vertaalwoordeboek opgeneem word? Tradisioneel kopieer leksikograwe in 'n groot mate die leksikale elemente van soortgelyke woordeboeke, maar Tarp (2004: 314) maan dat geen data ingesluit moet word omdat dit nog altyd so gedoen is of omdat ander bestaande woordeboeke dit het nie. Tog meen Roberts (1996: 18) dis 'n aanvaarbare punt om by te begin, maar dat alle finale besluite op die ontleding van geskikte korpusse gegrond moet word.

Tegnologiese ontwikkelings van die afgelope jare maak dit vir die eerste keer moontlik om op groot skaal verteenwoordigende elektroniese korpusse te bekom en tipiese patrone van werklike taalgebruik daarin te ontleed. Hierdie inligting kan sodoende die onderbou van woordeboeke vorm. Die BCD-projek het teen 1996 reeds 'n korpus van nagenoeg 310 miljoen woorde opgebou (Roberts 1996: 18); die COBUILD-argief bevat sowat 500 miljoen woorde (Sinclair et al. 1997: ix).

Die wesenskenmerke van 'n korpusgedrewe benadering is volgens Biber et al. (1998: 4) dat dit

- empiries is en werklike gebruikspatrone in natuurlike (outentieke) tekste ontleed;
- 'n groot versameling natuurlike tekste (die korpus) volgens bepaalde beginsels en metodes vir ontleding selekteer;
- die ontleding op die omvattende gebruik van rekenaars asook outomatiese sowel as interaktiewe tegnieke baseer (m.a.w. die leksikograaf kan in probleemgevalle sy/haar eie linguistiese oordeel gebruik); en
- sowel kwantitatiewe as kwalitatiewe ontledingstegnieke aanwend.

'n Korpusgedrewe benadering lei tot leksikografiese produkte wat verteenwoordigend en gebalanseerd is. Simpson (2002: 9) meen: "Nowadays, it would be a brave lexicographer who did not at least associate his or her dictionary with a corpus, however comprehensively that corpus was actually used to extract lexical information."

Biber et al. (1998: 23-24) se korpusgedrewe leksikografiese navorsing stel veral na ses belangrike vrae ondersoek in:

- Watter betekenis word met 'n bepaalde woord geassosieer?
- Hoe dikwels word 'n bepaalde woord relatief tot ander verwante woorde gebruik?
- Met watter nielinguistiese patrone (bv. registers, tydperke in die geskiedenis, streektaal) word 'n bepaalde woord geassosieer?
- Watter ander woorde word algemeen saam met 'n bepaalde woord gebruik, en hoe lyk die verspreiding van hierdie kollokasionele verbindings ("collocational sequences") oor verskillende registers heen?

- Hoe lyk die verspreiding van betekenisonderskeidings en gebruikstoepassings ("senses and uses") van 'n woord?
- Hoe verskil die gebruik en verspreiding van woorde wat skynbaar sinonieme is?

By die gebruik van Afrikaanse korpusse om "tipiese patrone van werklike taalgebruik" vas te stel, is 'n belangrike vraag of leksikograwe beskrywend of voorskrywend te werk moet gaan. Feinauer (2001: 165) wys op die "massiewe" invloed wat Engels toenemend op Afrikaans het weens die oorheersende rol wat Engels daaglik in die ervaringswêreld van elke individu speel. Dit bring tot so 'n mate 'n konseptuele verandering mee in die leksikalisering van konsepte in Afrikaans deur moedertaalsprekers, dat die taalgebruik in sommige van die voorste Afrikaanse koerante en radio- en televisieprogramme deurspek raak van leenvertalings en geleende taalvorme. "The original Afrikaans conceptualized form seems alien, too formal and prosaic to the everyday user of Afrikaans" (Feinauer 2001: 166). By die uitsoek van Afrikaanse korpusse en die selektering en ordening van leksikale materiaal moet hierdie tendens oorweging geniet.

Ten opsigte van die makrostruktuur en die dataverspreidingstruktuur is 'n belangrike besluit watter elemente hooflemmastatus moet kry. Afkortings? Morfeme? Akronieme? Gaan kollokasies en idiome aparte inskrywings wees of nis- en neslemmas vorm? Watter multileksikale items gaan as hooflemmas opgeneem word?

Nadat die span leksikograwe besluit het watter leksikonelemente in die vertaalwoordeboek opgeneem behoort te word, besluit hulle watter inligting oor elk van die lemmas verstrekkend gaan word. Hierdie bewerking van die lemmas verteenwoordig die mikrostruktuur, oftewel die interne struktuur van elke artikel. Hieraan word in meer besonderhede in paragrawe 5 en 6 hieronder aandag geskenk.

4. Die agtertekste-afdeling

Elemente wat vir opname in die agtertekste-afdeling van die voorgestelde vertaalwoordeboek oorweeg kan word, is die volgende:

- Omrekeningstabelle vir mate (lengte, oppervlak, volume, e.s.m.), gewigte en temperatuur ten einde voorsiening te maak vir verskille in Amerikaanse, Britse en Suid-Afrikaanse gebruike. (Die rede waarom hierdie soort inligting vir die vertaler in Afrikaans van nut kan wees, is omdat outeurs — bv. dosente wat lesingmateriaal voorberei met behulp van Amerikaanse of Britse handboeke — nie altyd vir hul Suid-Afrikaanse lesers die nodige aanpassings maak nie sodat die vertaler dit dan moet doen.)

- Kultuurgebonde data.
- Plekname asook beginsels vir die transliterasie van plekname in Afrikaans.
- Besonderhede van hulpbronne en kontakte wat vir vertalers nuttig kan wees, byvoorbeeld die name en/of internetadresse van organisasies en verenigings soos die Suid-Afrikaanse Vertalersinstituut, die English Academy of Southern Africa, die Departement van Justisie se Regsbiblioteek vir die name van wette, die Departement van Kuns en Kultuur se Nasionale Taal- en Terminologiediens, universiteite se vertaal- of taalskole, e.s.m.
- 'n Ekwivalentregister wat die woordeboek politoeganklik kan maak. (Omdat dit die gebruiksbestek van die woordeboek verhoog, kan dit 'n baie nuttige buiteteke wees.)

5. 'n Voorgestelde formaat vir die mikrostruktuur van 'n vertaalwoordeboek

Die voorgestelde formaat vir die bewerking van 'n aantal voorbeeldlemmas hieronder is naastenby op die formaat geskoei wat die BCD-projek gebruik (Roberts 1992a, Roberts 1992b, Roberts 1996 en die webwerf <http://www.dico.uottawa.ca/entries-en.htm>). Dis 'n tentatiewe formaat — die praktyk sal verdere verfyning vereis — wat soos volg daar uitsien:³

Inleidende afdeling

trefwoord met lettergreepindeling > woordsoort > fonetiese transkripsie > (spelvariant) > (verbuiging van werkwoorde)

Betekenisafdeling

(**nommer** van betekenisonderskeiding) > (leksikografiese etiket BT) > parafrase BT > parafrase/ekwivalente/sinonieme DT > (leksikografiese etiket DT) > voorbeeldmateriaal/vrye woordverbindings BT + DT > (kollokasies BT + DT) > (brontaalsinonieme)⁴ (kruisverwysings) > (brontaalantonieme)

(Afdeling vir vaste uitdrukkinge)

(**uitdr**) (BT + DT) > (kruisverwysing)

Afdeling vir samestellings

(**ss**) > (BT + DT) > (kruisverwysing)

(Die elemente tussen hakies verskyn nie noodwendig in elke artikel nie.)

Die voorbeeld wat die BCD bied, word om verskeie redes nuttig en relevant geag by die saamstel van 'n vertaalwoordeboek in Engels en Afrikaans:

- Dis 'n uitgesproke doelstelling van die BCD-projek om 'n woordeboek vir gesofistikeerde taalgebruikers, byvoorbeeld professionele vertalers, te ontwerp. Roberts, direkteur van die projek, is self 'n ervare professionele vertaler wat met ontwikkelings in vertaalstudie tred hou.
- Kanada is amptelik tweetalig en die taalwerkers in Frans en Engels het dieselfde gevorderde vlak van tweedetaalbeheer as taalwerkers in Afrikaans en Engels. Dit het direk betrekking op hul potensiële woordeboekgebruik.
- Gesofistikeerde woordeboekgebruikers het meer as basiese vertaalekwivalente nodig. Om daardie rede bevat die BCD meer inligting, en meer gegewens oor daardie inligting, as die meeste enkelvolumewoordeboeke: tegniese terme, multileksikale items (samestellings, vaste uitdrukkings en kollokasies), soveel doeltaalekwivalente as moontlik, betekenis-aanduiding by brontaallemmas, betekenisonderskeidings vir doeltaalekwivalente, verskillende voorbeelde van vrye verbindings wat die lemma kan vorm en verskeie moontlike vertalings daarvan (Roberts 1992a: 219).
- Die BCD dien as 'n voorbeeld van hoe al dié materiaal georden kan word sodat besige taalpraktisyns, soos vertalers wat gewoonlik onder tyddruk werk, dit vinnig en maklik kan naslaan.
- Die BCD gee die betekenisverklaring van lemmas in die brontaal. Navorsing oor die tussentalige vermoëns van vertalers (Presas 2000: 21-26) toon dat dit 'n gekoördineerde tweetalige taalgeheue bevorder, waarvoor professionele vertalers ideaal gesproke moet beskik, en ook die vertaler se skeppende vermoë tot diens kan wees (Kussmaul 1995: 51).
- Die baanbrekerswerk van die BCD is goed gedokumenteer en die metodes wat die samestellers gevolg het, is daarom vir ander woordeboeksamestellers toeganklik.

'n Voorbeeld van hoe die BCD trefwoorde hanteer, werp lig op die struktuur waarvolgens dié woordeboek inligting bewerk en aanbied (*Bilingual Canadian Dictionary: Sample Entries* 2004):

car pool *n* or **carpool** or **car-pool**

1 (*arrangement whereby several commuters travel together in one car and share costs*) covoiturage *m*. ***I commute to work by car pool** je fais du covoiturage pour aller au travail.

2 (*group of commuters travelling together in such an arrangement*) groupe de covoiturage *m*. ***to belong to a car pool** faire partie d'un groupe de covoiturage.

3 (*number of cars owned by a company or organization for the use of its members*) parc automobile *m*.

(**cmp1**) **car pool lane** *n* voie réservée au covoiturage *nf*; **car pool lot** *n* stationnement réservé aux covoitureurs *nm* [(CD)].

carpool *vi* or **car pool** or **car-pool**

1 (*to travel with other commuters in one car and share costs*) faire du covoiturage,

covoiturer [(CD)]. *seven of us carpool to work nous sommes sept à faire du covoiturage pour aller au travail = nous sommes sept à covoiturer pour aller au travail.

carpool vt or **car pool** or **car-pool**

1 (to take turns driving passengers somewhere) expl.: conduire des passagers quelque part. *it is my turn to carpool the neighbourhood's kids to school this week c'est à mon tour de conduire les enfants du voisinage à l'école cette semaine.

car pooler n or **carpooler** or **car-pooler**

1 (member of a group of commuters who travel together in one car and share costs) covoituteur m [(CD)], f covoitureuse [(CD)]. *car poolers are allowed to park here all day il est permis aux covoitureurs de stationner ici toute la journée.

car pooling n or **car-pooling** or **carpooling**

1 (transportation in a private vehicle of several passengers who share operating costs) covoiturage m. *car pooling reduces traffic congestion and air pollution le covoiturage réduit les embouteillages et la pollution de l'air.

By die konstruksie hieronder van woordeboekinligting vir die voorgestelde Engels–Afrikaans-vertaalwoordeboek is dimensies bygevoeg wat die BCD nie het nie. Eerstens word die lettergreepindeling en verbuigde vorme van die lemma aangetoon, omdat dit by die versorging van dokumente vir taalwerkers belangrik is. Tweedens word sinonieme van die lemma waar moontlik gelys, omdat dit ingevolge beskrywings van die vertaalproses (sien veral Presas 2000: 21-25; Neubert en Shreve 1992: 43-46; Mossop 1998: 40; Kussmaul 1995: 40-51; Tarp 2002: 73 en Crafford 2005: 56-59) kan meehelp dat die vertaler 'n gepaste vertaling maak. Die brontaalsinonieme ondersteun twee kommunikasiegerigte leksikografiese funksies van die vertaalwoordeboek, te wete om vertalers met die begrip van tekste in hul tweede taal en die vertaling van tekste uit hul tweede taal in hul eerste taal te help.

Die leksikale items wat hieronder bespreek word, is só gekies dat die voorgestelde formaat met betrekking tot verskeie sake getoets kan word. Die bewerking van die leksikale items neem van paragraaf 6.1 tot 6.2 in kompleksiteit toe. Eerste aan die beurt is 'n leksikale item (**incur**) met 'n enkele woordsoortelike funksie en slegs een betekenisonderskeiding, maar wat 'n aantal kollokasionele verbindings kan vorm. Daarna volg 'n leksikale item (**endeavour**) met meer as een woordsoortelike funksie, maar slegs een betekenisonderskeiding vir elke funksie. Dié item illustreer die hantering van 'n wisselvorm wat spelwyse betref, die gebruik van leksikografiese etikette, die optrede van die item in 'n vaste kombinasie met 'n voorsetsel, 'n samestelling waar die eerste element nie die lemma is nie, en 'n kruisverwysing. Die volgende leksikale item (**juggle**) is veral gekies omdat Afrikaanssprekendes die Engelse woord dikwels net so leen. Dié item illustreer onder meer die gebruik van leksikografiese etikette en die hantering van 'n vaste uitdrukking.

Hierna kom twee multileksikale leksikale items aan die beurt. Die eerste (die lemma **big bang theory**) is 'n leksikale item wat net een woordsoortelike funksie het en wat die gebruik van 'n leksikografiese etiket en die aanbied van ensiklopediese inligting toon. Die multileksikale leksikale item **fast track** het meer as een woordsoortelike funksie en toon die hantering van 'n wisselende

spelvorm by funksiewisseling, asook die hantering van 'n idiomatiese uitdrukking.

Die volgende woordeboeke (in alfabetiese volgorde) is ruimskoots benut om leksikale inligting vir die bewerkingsvoorstelle hieronder te bekom.

- *Collins COBUILD English Dictionary*, 1997
- *Collins Gem Dictionary of English Spelling*, 1994
- *Longman Business English Dictionary*, 2001
- *Pharos New Words/Nuwe Woorde*, 1999
- *The Reader's Digest Oxford Complete Wordfinder*, 1992
- *South African Concise Oxford Dictionary*, 2002
- *Tweetalige Woordeboek/Bilingual Dictionary*, 1984
- *Verklarende Handwoordeboek van die Afrikaanse Taal*, 2000

6. 'n Praktiese model vir 'n vertaalwoordeboek

Dis belangrik om daarop te let dat woordeboekmaak nie 'n taak is wat deur enkelinge verrig word nie. In 'n werklike situasie sal alle besluite soos dié hieronder aan die oordeel van 'n paneel van kundiges onderworpe wees. Hierbenewens behoort die beginsel van gelyktydige terugvoer (kyk paragraaf 7 hieronder) van meet af aan en deurlopend toegepas te word ten einde die voorgestelde vertaalwoordeboek deur middel van formatiewe evaluering tot stand te bring.

Voorstelle vir die bewerking van 'n aantal lemmas word vervolgens aangebied, vergesel van annotasies wat redes vir besluite uiteensit. Hierdie besluite is deurgaans daarop gemik om die verskillende leksikografiese funksies van 'n vertaalwoordeboek (Crafford 2005: 54-59) te ondersteun.

6.1 'n Leksikale item met 'n enkele woordsoortelike funksie en slegs een betekenisonderskeiding

in-cur vt /ɪŋkɜː/ **in-curs, in-cur-ring, in-curred**

(become subject to something [unwelcome or unpleasant] as a result of one's actions) jou blootstel aan iets (onaangenaams) of jou iets (onaangenaams) op die hals haal * **he behaves badly and incurs the wrath of the community** hy gedra hom sleg en haal hom die gramskap / woede van die gemeenskap op die hals.

~ **blame** skuld kry; ~ **costs / debts / expenses** (on)koste / skuld(e) / uitgawes aangaan / maak; ~ **danger / risk** gevaar loop, jou aan gevaar blootstel; ~ **displeasure / wrath** jou onguns / onmin / woede op die hals haal; ~ **a fine / liability / loss** 'n boete / aanspreeklikheid / verlies op die lyf loop; ~ **an illness / injury** 'n siekte / besering opdoen; ~ **a penalty / punishment** boete / straf verdien / ontvang.

— SYNONYMS *attract, arouse, bring upon or on (oneself), come in for, draw, experience, expose (oneself) to, face, invite, lay (oneself) open to, meet (with), provoke, suffer, sustain, undergo*

Verklarende aantekeninge en rasionalisering van besluite

vt	Woordsoortaanduiding — die <i>v</i> staan vir werkwoord (verb) en die <i>t</i> dui aan dat dit 'n oorganklike (<i>transitive</i>) werkwoord is (in die voortekstefdeling moet alle redaksionele afkortings verklar word). Oorganklikheid of onoorganklikheid is soms bepalend vir besluite oor 'n vertaling.
/ɪŋkɛ:/ (Die fonetiese alfabet wat hier gebruik word, is gebaseer op die simbole van die Internasionale Fonetiese Alfabet.)	Uitspraak en klem word aangedui. Uitspraakleiding is belangrik. ⁵ As taalpraktisyns strek vertalers se take en pligte wyer as die geskrewe woord en dit is derhalwe belangrik vir hulle om die korrekte uitspraak van woorde te ken.
in-curs, in-cur-ring, in-curred	Verbuigings van werkwoorde word in die inleidende deel gegee om naslaan te vergemaklik. Woordafbreking word aangedui omdat dit belangrik is by die proeflees en taalversorging van dokumente.
Geen etimologiese inligting nie	Dis debatteerbaar of 'n vertaalwoordeboek etimologiese inligting moet bevat. Sommige vertalers meen dit moet wel gegee word, want afgesien van taalpraktisyns se belangstelling in woorde en hul oorsprong, kan etimologiese kennis hulle help om skeppend te vertaal. ⁶ Die woord palimpse se Griekse etimologie [G. <i>palin</i> weer + <i>psao</i> skoon vryf] kan die vertaler byvoorbeeld herinner aan die uitdrukking "om 'n skoon lei te hê".
<i>become subject to (something unwelcome or unpleasant) as a result of one's actions</i>	Die betekenisverklaring in die brontaal geskied d.m.v. 'n kort omskrywing of parafrase.
jou blootstel aan iets (onaangenaams) of jou iets (onaangenaams) op die hals haal	Die betekenisverklaring vind plaas aan die hand van 'n doeltaalparafrase en/of doeltaalsinonieme.
* he behaves badly and incurs the wrath of the community	Voorbeeldmateriaal dra sintaktiese en sintagmatiese inligting op 'n implisiete wyse oor (dis veral nuttig by vertaling in T2). Voorbeeldmateriaal moet verkieslik op korpusnavorsing in parallelle tekste gegrond wees en nie deur die leksikograaf self geskep word nie. ⁷
hy gedra hom sleg en haal hom die gramskap / woede van die gemeenskap op die hals	Die professionele vertalers in die leksikografiese span vertaal voorbeeldmateriaal, indien nodig m.b.v. parallelle tekste.

<p>~ <i>blame</i> skuld kry; ~ <i>costs / debts / expenses</i> (on)koste / skuld(e) / uitgawes aangaan / maak; ~ <i>danger / risk</i> gevaar loop, jou aan gevaar blootstel; ~ <i>displeasure</i> jou onguns / onmin op die hals haal; ~ <i>a fine / liability / loss</i> 'n boete / aanspreeklikheid / verlies op die lyf loop; ~ <i>an illness / injury</i> 'n siekte / besering opdoen; ~ <i>a penalty / punishment</i> boete / straf verdien / ontvang.</p>	<p>Kollokasies en voorgestelde vertaalekwivalente, is alfabeties georden. (Ordering op grond van gebruiksfrekwensie moet op korpusnavorsing berus.)</p>
<p><i>arouse, attract, bring upon or on (oneself), come in for, draw, experience, expose (oneself) to, face, invite, lay (oneself) open to, meet (with), provoke, suffer, sustain, undergo</i></p>	<p>Absolute en gedeeltelike sinonieme in die brontaal word gegee om die volle betekenispotensiaal van die lemma incur aan te dui. Waar toepaslik moet antonieme om dieselfde rede gelys word. Hulle is alfabeties georden. Ordering op grond van gebruiksfrekwensie moet op korpusnavorsing berus.</p>

6.2 'n Leksikale item met meer as een woordsoortelike funksie en een betekenisonderskeiding elk

en.deav.our vt /ɪndɛvə/ (US endeavor) **en.deav.ours, en.deav.our.ing, en.deav.oured** (try hard [to do or achieve something]) 'n poging aanwend, poog (*deftig*), probeer, streef, trag (*deftig*), ywer, of jou beywer (om iets te doen of te bereik) * **I will endeavour to arrange a meeting** ek sal poog (*deftig*) om 'n ontmoeting te reël = ek sal 'n ontmoeting probeer reël
— SYNONYMS *aim, aspire, attempt, do one's best, essay (formal), exert oneself, have a go or crack or whack or shot at (colloq.), make an effort, strive, struggle, take a stab at (colloq.), try*

en-deav-our n /ɪndɛvə/ (US endeavor) (*act of endeavouring, an earnest and industrious effort, an enterprise — [a formal use]*) poging, strewe, stryd, onderneming, waagstuk * **It is my constant endeavour to ...** Ek streef steeds daarna om ...
— SYNONYMS *attempt, effort, enterprise, essay (formal), pains, striving, struggle, undertaking, venture*
(*cmp*) **scientific endeavour** n wetenskapsbeoefening (→ **scientific**)

Verklarende aantekeninge en rasionalisering van besluite

<p>(US endeavor)</p>	<p>Verskeie Engelse woorde het twee of meer spelwyses. By die produksie of kontrolering van geskrewe tekste moet die vertaler hiervan kennis dra. In die voortekste-afdeling sal daar 'n vermelding wees dat wisselvorme gegee word.</p>
----------------------	--

Leksikografiese etikette ('formal', 'colloq.', 'deftig')	Pragmatiese inligting m.b.v. gebruiksetikette help vertalers met sowel die resepsie as die produksie van tekste.
* I will endeavour to ...	In die voorbeeldsin word twee punte implisiet tuisgebring: (a) die gebruik van <i>will</i> plaas van <i>shall</i> om vaste voorneme oor te dra en (b) die feit dat <i>endeavour</i> met die voorsetsel <i>to</i> verbind.
cmp ... (→ scientific)	By 'n samestelling (<i>compound</i>) waarvan die eerste element nie die lemma is nie, verskyn 'n kruisverwysing (aangedui m.b.v. 'n pyltjie) na die plek waar dié element as lemma optree.

6.3 'n Leksikale item met meer as een woordsoortelike funksie en meervoudige betekenisonderskeidings

jug-gle vt, vi /dʒʌgəl/ **jug-gles, jug-gling, jug-gled**

1 (*continuously toss in the air and catch a number of objects so as to keep at least one in the air at any time*) balle in die lug hou, jongleer (*deftig*, uitspr. / ʒɔŋlɪər /) * **to juggle with something** met iets jongleer (balle, ens.)

2 (*cope with by adroitly balancing [several activities]*) baie dinge gelyk doen / hanteer / inpas; behendig / knap werk; oral raakvat
— SYNONYMS *manipulate, manoeuvre*

3a (*rearrange adroitly*) goël, goëlkunsies uitvoer, rats herskik, regsommel, toor
— SYNONYMS *alter, change, modify, reposition, reset, reshuffle, shift, shuffle, switch, transpose*

3b (*deceive or cheat*) bedrieg, fop, kul
— SYNONYMS → **deceive**

3c (*misrepresent [facts]*) (feite) verdraai, verwring, foutief voorstel, (*infrm.*) dokter, kook
— SYNONYMS *disguise, distort, falsify, massage, tamper with, rig, fix or doctor or cook or fiddle* (colloq.)

(*exp*) **juggle someone out of something** iets van iemand afrokkel

jug-gle n

1 (*a piece of juggling*) baltoertjie, goëltoertjie

— SYNONYMS *dexterity, wizardry, sleight of hand*

2 (*a fraud*) bedrieëry, kullery, oëverblindery, skelmstreek

— SYNONYMS *deceit, deception, trickery, swindle, sophistry, hocus-pocus, hoax, trick, funny business or rip-off or scam* (colloq.)

Verklarende aantekeninge en rasionalisering van besluite

vt, vi	Dié aanduiding toon dat die werkwoord beide oorganklik en onoorganklik gebruik kan word.
--------	--

<p>balle in die lug hou, jongleer (<i>def-tig</i>, uitspr. / ʒɔŋlɪər /)</p> <p>* to juggle with something met iets jongleer (balle, ens.)</p>	<p>Die woord <i>jongleer</i> is in Nederlands taamlik bekend, maar in Afrikaans is dit weinig gebruiklik. Dit weerspieël 'n toenemende leksikale gaping. Omdat sommige vertalers d.m.v. terugvoer op ZaLang ('n elektroniese gespreksgroep vir vertalers) egter aangedui het dat hulle die term ken en selfs gebruik, is besluit om dit wel op te neem, met 'n aanduiding van uitspraak. (Uitspraak in die doeltaal — m.b.v. dieselfde fonetiese konvensie as vir die brontaal — hoef slegs by uitsondering en dan veral by vreemde woorde gegee te word.)</p>
<p>3b SYNONYMS → deceive</p>	<p>Daar is 'n kruisverwysing by 3b omdat dit duplisering sal wees om die sinonieme vir 'deceive' hier te lys.</p>
<p>(Betekeinsonderskeidings 1-3c)</p> <p>3c verdraai, verwring, foutief voorstel, (<i>infml.</i>) dokter, kook</p>	<p>Betekeinsonderskeidings word genommer en apart gehanteer ter wille van duidelikheid. Ekwivalente word vir elke betekeinsonderskeiding aangedui. Die leksikale gaping word oorbrug m.b.v. parafrasering. Die gebruiksetiket <i>infml.</i> dui informele register aan.</p>
<p>(<i>exp</i>) juggle someone out of something iets van iemand afrokkel</p>	<p>Vaste uitdrukings (<i>fixed expressions</i>) word apart gelys om dit vir die gebruiker makliker te maak om hulle te vind.</p>

Afleidings soos **juggler**, **jugglery** en **juggling act** vorm nie deel van die bewering van die lemma **juggle** nie, maar sal as aparte lemmas opgeneem word.

6.4 'n Multileksikale leksikale item met slegs een woordsoortelike funksie

big bang theory n

(astronomy) (*theory that suggests that the origin of the universe was marked by the rapid expansion of matter from a state of extremely high density and temperature*) oerknalteorie, teorie dat 'n ontploffing tot die ontstaan van die heelal gelei het

— SYNONYMS *big bang theory* = *superdense theory*

Verklarende aantekeninge en rasionalisering van besluite

<p>big bang theory</p>	<p>Die uitspraak word nie aangedui nie, omdat dit by die lemmas big, bang en theory gedek sal word. By uitdrukings in 'n ander taal, soos faux pas of sine qua non, is uitspraakleiding essensieel.</p>
<p>(astronomy)</p>	<p>Vaketiket.</p>

(theory in astronomy that suggests ...)	Die vertaalwoordeboek vervul 'n kennisfunksie deur onder meer ensiklopediese inligting in die brontaal te verskaf. Dit word in maklik verstaanbare eerder as wetenskaplike taal gegee. ⁸
---	---

6.5 'n Multileksikale leksikale item met meer as een woordsoortelike funksie

fast track n /fɑ:stɹæk/

(a rapid route or method [to reach a goal]) (fig.) kortpad (*na bo*); dolle vaart (*na nêrens*)

* **put something on the fast track** die pas van iets versnel

(*exp*) **live on the fast track** 'n vinnige pas handhaaf

fast-track v

(accelerate the development or progress of) die pas van iets versnel, iets bespoedig / deurjaag

* **his career was fast-tracked in the interest of affirmative action** sy loopbaanvoordringing is bespoedig ter wille van regstellende aksie

Verklarende aantekeninge en rasionalisering van besluite

fast-track v	Die gebruiker sal oplet dat die werkwoord met 'n koppelteken gespél word.
---------------------	---

7. Die konsep van gelyktydige terugvoer

Volgens De Schryver en Prinsloo (2000: 3) is dit "algemeen gebruiklik" dat woordeboekgebruikers aan 'n reeks toetse onderwerp word ten einde hul sukses wat betref die inwin van inligting te meet, m.a.w. te bepaal hoe suksesvol 'n voltooide woordeboek sy funksie vervul. Gewoonlik kan sodanige terugvoer eers in berekening gebring word by 'n volgende of hersiene uitgawe van die gepubliseerde woordeboek. Ideaal gesproke moet teikengebruikers egter terugvoer kan verskaf terwyl die samestelling van die woordeboek nog aan die gang is. Die proses van gelyktydige terugvoer maak hierdie ideaal haalbaar. Dis 'n proses waardeur teikengebruikers se terugvoer verkry word "both from the very start and during the entire compilation process" (De Schryver en Prinsloo 2000: 4). Dit geskied met behulp van 'n reeks bekostigbare, kleinskaalse woordeboekprojekte wat aanvullend van stapel gestuur word en parallel aan die hoofprojek loop. Die kleiner, parallelle projekte het dieselfde basiese struktuur, inhoud en teikengebruikers as die hoofprojek, maar word oor 'n aantal korter periodes afgehandel. Vanaf die oomblik wat die eerste parallelle projek geloods word, word informele en formele terugvoer ingewin en die inligting in die hoofprojek terugvoer. So word die parallelle projekte deurlopend as eksperimentele instrumente gebruik om verskeie strategieë te beproef en die aanbied

van inligting in die hoofprojek te verfyn. Die proses kom slegs tot 'n einde wanneer die hoofprojek afgehandel is.

Die gebruik van nuwer tegnologie, soos die internet, kan deesdae 'n dinamiese proses van woordeboekverbetering moontlik maak deurdat dit leksikografe in staat stel om deurlopend van gebruikers terugvoer te ontvang.

8. Ten slotte

De Schryver en Prinsloo (2000: 3) meen dis belangrik dat woordeboeksamstellers nie slegs hul teikengebruikers se behoeftes nie, maar ook hul wense in ag moet neem. Dit maak van teikengebruikers die "bevoorregte kritici" van 'n nuwe woordeboek. Vertalers self moet dus besluit wat hulle in 'n vertaalwoordeboek wil hê en hoe hulle dit wil hê, want volgens die beginsel van gelyktydige terugvoer besluit gebruikers watter data ingesluit of uitgelaat en op watter wyse dit aangebied moet word.

Notas

1. Die BCD-projek is 'n langtermyn- interuniversitêre projek wat daarop gerig is om 'n tweetalige woordeboek tot stand te bring wat taalgebruik in Kanadese Engels en Kanadese Frans weerspieël (Roberts 1996: 10).
2. In sommige tale bestaan die moontlikheid van 'n meer omvattende toegangsalfabet wat dan implikasies vir die aantal artikelreeks en die ordening daarvan het.
3. Die afkortings BT en DT staan vir "brontaal" en "doeltaal" onderskeidelik, terwyl uitdr. "vaste uitdrukkings" en ss "samestellings" aandui. Sowel die brontaal as die doeltaal kan as metataal gebruik word, afhangende van die onmiddellike taalomgewing.
4. Die gebruiker sal in die voortekste-afdeling ingelig word dat sowel absolute as gedeeltelike sinonieme alfabeties gelys word en dat die sinonieme mekaar nie oral kan vervang nie.
5. Roberts (1996: 22) stem nie saam nie. Sy meen uitspraak "is the aspect of form that is least important for translators, who have to know how to write correctly but not necessarily how to pronounce correctly".
6. Roberts (1996: 22-23) meen egter etimologiese inligting kan vertalers mislei en selfs aanleiding gee tot die verskynsel van vals vriende. Volgens haar kan sommige woorde in 'n vertaler se twee werkstale dieselfde oorsprong hê en daarom vormlik ooreenstem, maar kon hulle mettertyd 'n effense verskil in betekenis ontwikkel het.
7. Die internet maak voorbeeldmateriaal in woordeboeke toenemend oorbodig vir vertalers wat toegang het tot die wêreldwye web en daar kan opsoek hoe woorde in bepaalde kontekste optree.
8. De Beaugrande (2003: 3) meen dat die taak om definisies te verskaf in die verlede te dikwels oorgelaat is aan spesialiste in verskillende vakrigtings, wat dan met definisies "no less obscure than the expression itself" vorendag gekom het. Die gebruik van korpusse kan verseker dat die betekenis wat gegee word werklik in gewone tekste voorkom, wat nie noodwendig is hoe 'n spesialis dit sou stel nie.

Bronnelys

Woordeboeke

- Bilingual Canadian Dictionary: Sample Entries*. 2004. <http://www.dico.uottawa.ca/entries-en.htm> [2 November 2004].
- Bosman, D.B., I.W. van der Merwe en L.W. Hiemstra**. 1984^s. *Tweetalige Woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg.
- Collins Gem Dictionary of English Spelling*. 1994. Londen: HarperCollins.
- Du Plessis, Madaleine**. 1999. *New Words and Previously Overlooked Ones/Nuwe woorde en oues wat in die slag gebly het. English-Afrikaans*. Kaapstad: Pharos.
- Kavanagh, K., D. Mantzel, T. van Niekerk, J. Wolvaardt en M. Wright**. 2002. *South African Concise Oxford Dictionary*. Kaapstad: Oxford University Press.
- Longman Business English Dictionary*. 2001. Harlow: Pearson Education.
- Odendal, F.F., P.C. Schoonees, C.J. Swanepoel, S.J. du Toit en C.M. Booysen**. 2000^s. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Johannesburg: Perskor.
- Sinclair, J., G. Fox en S. Bullon (Reds.)**. 1997. *Collins COBUILD English Dictionary*. Londen: HarperCollins.
- Tulloch, Sara**. 1993. *The Reader's Digest Oxford Complete Wordfinder*. Londen: The Reader's Digest Association.

Ander

- Biber, Douglas, Susan Conrad en Randi Reppen**. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Crafford, M.F.** 2005. *Vertalers en hul bronne: Die behoefte aan 'n vertaalwoordeboek met Engels en Afrikaans as behandelde taalpaar*. Ongepubliseerde M.Phil.-skripsie. Stellenbosch: Universiteit van Stellenbosch.
- De Beaugrande, Robert**. 2003. *Function and Form of Reference Works: Setting a Record Straight for J. McH. Sinclair and the COBUILD*. <http://beaugrande.com/SinclairCarstensen.htm> [27 Februarie 2003].
- De Schryver, Gilles-Maurice en D.J. Prinsloo**. 2000. The Concept of 'Simultaneous Feedback': Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.
- Feinauer, Ilse**. 2001. Language Contact between Afrikaans and English Leads to Conceptual Change in Afrikaans. Thelen, M. en B. Lewandowska-Tomasczcyk (Reds.). 2001. *Translation and Meaning Part 5*: 163-171. Maastricht: Hogeschool Zuyd, Maastricht School of International Communication.
- Gouws, Rufus H.** 2001. Lexicographic Training: Approaches and Topics. Emejulu, J. Duplessis (Red.). 2001. *Éléments de lexicographie gabonaise, Tome I*: 58-94. New York: Jimacs-Hillman.
- Keyser, Helane**. 2003. Ondersoek na 'n model vir die opleiding van leksikograwe vir verklarende woordeboeke. Ongepubliseerde M.Phil.-skripsie. Stellenbosch: Universiteit van Stellenbosch.
- Kussmaul, Paul**. 1995. *Training the Translator*. Amsterdam/Philadelphia: John Benjamins.
- Louw, Phillip A.** 1999. Access Structures in a Standard Translation Dictionary. *Lexikos* 9: 108-118.

- Mossop, Brian.** 1998. The Workplace Procedures of Professional Translators. Chesterman, A., N. Gallardo San Salvador en Y. Gambier (Reds.). 1998. *Translation in Context: Selected Contributions from the EST Congress, Granada 1998*: 39-48. Amsterdam: John Benjamins.
- Neubert, Albrecht en Gregory M. Shreve.** 1992. *Translation as Text*. Kent, Ohio: Kent State University.
- Presas, Marisa.** 2000. Bilingual Competence and Translation Competence. Schäffner, C. en B. Adab (Reds.). 2000. *Developing Translation Competence*: 19-31. Amsterdam/Philadelphia: John Benjamins.
- Roberts, Roda P.** 1990. Translation and the Bilingual Dictionary. *Meta* 25(1): 74-82.
- Roberts, Roda P.** 1992a. Organization of Information in a Bilingual Dictionary Entry. *EURALEX '90 Proceedings. IV International Congress*: 219-231. Barcelona: Biblograf in samewerking met die Kommissie vir Europese Gemeenskappe.
- Roberts, Roda P.** 1992b. Methods of Bilingual Dictionary-making: The Canadian Experience. Hyldgaard-Jensen, K. en A. Zettersten (Reds.). 1992. *Symposium on Lexicography V*: 91-116. Tübingen: Max Niemeyer.
- Roberts, Roda P.** 1996. Dictionaries, Dictionary-making and Translation. Nistal, P.F. and J.M.B. Gozalo (Reds.). 1996. *A Spectrum of Translation Studies*: 9-35. Valladolid: Universidad de Valladolid, Servicio de Apoyo a la Enseñanza.
- Simpson, John.** 2002. The Revolution of English Lexicography. *Dictionaries: The Journal of the Dictionary Society of North America* 23: 1-15.
- Tarp, Sven.** 2002. Translation Dictionaries and Bilingual Dictionaries: Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Tarp, Sven.** 2004. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.

A New Way to Lemmatize Adjectives in a User-friendly Zulu–English Dictionary

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

Abstract: Traditionally, Zulu adjectives have been lemmatized under their stems only. In this research article, an in-depth analysis is undertaken to make a case for the lemmatization of all frequent adjectival forms *with their adjective concords* rather. It is shown that the supposed explosion in size of the dictionary may be contained within a corpus-driven Sinclairian framework. The advantages of such a word-like treatment far outnumber the generalizations that have hitherto characterized the lexicographic treatment of adjectives in Zulu. The study is supported by ample dictionary extracts from a Zulu–English dictionary project aimed at junior users. Comparisons with existing dictionaries and textbook data are also made.

Keywords: LEXICOGRAPHY, LINGUISTICS, GRAMMAR, DICTIONARY, BILINGUAL, CORPUS, LEMMATIZATION, FREQUENCY, ZULU (ISIZULU), ENGLISH, ADJECTIVE, ADJECTIVE STEM, QUALIFICATIVE ADJECTIVE, COPULATIVE ADJECTIVE, USER-FRIENDLY, REAL EXAMPLE, COLLOCATION, COMBINATION, DERIVATION, IDIOMATIC USE, SEMANTIC PROSODY

Samenvatting: Een nieuwe manier om adjectieven te lemmatiseren in een gebruiksvriendelijk Zoeloe–Engels woordenboek. Traditioneel worden adjectieven in Zoeloe enkel onder hun stam gelemmatiseerd. In dit onderzoeksartikel wordt een grondige analyse uitgevoerd met het oog op de invoering van een nieuwe methode waarbij alle frequente adjectieven *met hun adjectiefschakel* in het woordenboek worden geplaatst. Er wordt aangetoond dat de vooronderstelde explosie in grootte van het woordenboek beperkt kan worden binnen een corpusgedreven Sinclairiaans kader. De voordelen van zo een woordachtige behandeling overstijgen ruimschoots de veralgemeningen die totnogtoe de lexicografische behandeling van adjectieven in Zoeloe hebben gekarakteriseerd. De studie wordt ondersteund door een groot aantal passages uit een Zoeloe–Engels woordenboekproject gericht op jonge gebruikers. Vergelijkingen met bestaande woordenboeken, alsook handboeken worden ook gemaakt.

Sleutelwoorden: LEXICOGRAFIE, LINGUÏSTIEK, GRAMMATICA, WOORDENBOEK, TWEETALIG, CORPUS, LEMMATISATIE, FREQUENTIE, ZOELOE, ENGELS, ADJECTIEF, ADJECTIEF STAM, KWALIFICEREND ADJECTIEF, COPULATIEF ADJECTIEF, GEBRUIKSVRIENDELIJK, ECHT VOORBEELD, COLLOCATIE, COMBINATIE, AFLEIDING, IDIOMATISCH GEBRUIK, SEMANTISCHE PROSODIE

1. From Bloomfield to Sinclair via Doke

Half a century ago, two excellent dictionaries for Zulu appeared, viz. Doke and Vilakazi's (1953) *Zulu-English Dictionary*, and Doke, Malcolm and Sikakana's (1958) *English-Zulu Dictionary*. The coverage, detail and meticulousness of these two dictionaries are of such a high standard that they had the ironic effect of stalling all future lexicographic efforts for Zulu. Indeed, to this date not a single dictionary for Zulu — whether bilingual or monolingual — has been compiled that comes even close to the quality of Doke's pair of dictionaries. Doke's pair remains the standard against which all current Zulu dictionaries are compared, and will likely remain the standard for many years to come.

In Doke and Vilakazi's Zulu to English dictionary, the so-called 'stem approach' to lemmatization is used, meaning that (a section of) the Zulu lexicon is grouped around word stems. The multitude of (often stacked) prefixes, suffixes and circumfixes which characterize a conjunctively written language such as Zulu have thus been cut off, with (supposed) meanings assigned to the resulting (extracted) stems. For linguists such an approach is arguably a magnificent and efficient lemmatization approach; for the average user it is problematic.

For about a decade now, we have informally observed the use of this Zulu dictionary at university level as well as within different language services of various government departments. We have noticed that, on average, as many as two look-up procedures are required before a user also finds what he/she is looking for. The main reason for this is not so much the result of inconsistencies in the lemmatization proper, but simply because a large amount of grammatical knowledge is presupposed before one can successfully consult this dictionary. This is valid for both decoding (receptive) and encoding (active) use, and for learners as well as mother-tongue speakers. Two random, straightforward examples follow to illustrate these points.

Zulu nouns in the gender 9/10 have the noun class prefixes *iN-* for the singular (class 9), and *iziN-* for the corresponding plural (class 10) — with *N* a nasal, i.e. *n* or *m*. A user of a stem-based dictionary may conclude that 9/10 nouns are lemmatized under the nasal *N*. So when wishing to look up, say, *indlovu/izindlovu* 'elephant/elephants' this user will go to the alphabetic stretch **N**. In this case, however, these words cannot be found there, as Doke realized that the stem here is not *-ndlovu*, but rather *-dlovu*, calling in the Ur-Bantu form of this noun stem (*-ɣoɣu*) to substantiate this. Neither learners nor mother-tongue speakers, however, can be expected to be versed in comparative or historical Bantu linguistics, so the finer points of Doke's lemmatization approach are entirely lost on all but a few of the most ardent users.

As an example to illustrate the encoding use of a Zulu dictionary, consider the ordinal 'fourth'. When used neutrally (as in 'she came fourth'), the form is *isine*; while a possessive concord needs to be prefixed to this form for definite uses (as in 'the fourth quarter'), resulting in forms such as *yesine*, *wesine*, *lesine*,

sesine, etc. In Doke, one needs to look up all these forms under *-ne* (the reasoning being that these forms are derived from the adjective stem *-ne* 'four'), but under *-ne* the differing ordinal uses (neutral vs. definite) are not stated explicitly. Linguists, of course, will see nothing wrong with this, as they will refer the dictionary user to the grammar for the actual *use*.

One solution is indeed to dissociate the grammar from the lexicon, recalling Bloomfield (1933: 274): 'The lexicon is really an appendix of the grammar.' At this point one could focus on, say, just nouns and verbs in a dictionary, and relegate all other word classes to the grammar. If this sounds too far-fetched, consider the latest monolingual dictionary for Zulu, *Isichazamazwi sesiZulu* (Mbatha 2006). In this dictionary's front matter, one reads that (a) only content words belong in a dictionary, and that (b) this means only four word classes are recognized in Zulu: noun, verb, exclamation or interjection, and ideophone. Probably realizing that this proposition is untenable, the compilers somehow 'forced' meanings onto extremely low-frequency to non-existing verb and noun stems. As such, one for instance finds the noun *í(li)nîngi* 'the majority' but not the adjective stem *-ningi* 'much/many'. Likewise, the extremely-low-frequency noun *împêla* 'the real one' — which is mostly used in possessive constructions, at which point it is a possessive — is found instead of the highly-frequency adverb *impela* 'really'.¹

Even though there are days on which the prospect surfaces to 'get rid of' all lemmatization and presentation problems in Bantu lexicography by this means, it is exactly the lexicographer's task not to give in here. Indeed, no sooner has one finished contemplating Bloomfield than Sinclair (1966: 422-423) must be considered:

We speak casually about 'fully grammatical items' or 'function words' as if there were items which were entirely irrelevant in the study of lexis. ... Every morpheme in a text must be described both grammatically and lexically ... Each successive form in a text is a lexical item or part of one, and there are no gaps where only grammar is to be found.

2. A user-friendly Zulu dictionary: mission statement

Against the background sketched in Section 1, a new type of (bilingual) Zulu dictionary has been envisaged, one which would also and for the first time be pitched at the level of junior users. The mission statement for this project has been described by De Schryver and Wilkes (2008: 831) as follows:

An approach which cuts down to the smallest morpheme level (as in Doke & Vilakazi) is user-unfriendly for the target user group envisaged, while an approach which throws out most word categories, and forces so-called core Zulu meanings onto the remaining section (as in Mbatha) is even more user-unfriendly. While the former is linguistically sound, the latter moreover is not.

The user-friendly approach/solution advocated here revolves around two notions: (a) except for verbs and a few exceptions (such as the conjunction *-thi*

(when), which behaves like a verb), all items from all word classes can be lemmatised *with their primary prefix(es)* included, as well as *with their suffixes* included; (b) overall *corpus frequencies* may be used in order to make a decision on the number of prefixes as well as which prefixes to include for each word class as a whole, and thus on how to organise/lemmatise the lexicon.

Implicit in this mission statement is that one has access to a large Zulu corpus, that one has a procedure to lemmatize this corpus (while keeping track of all individual as well as summed and overall corpus frequencies), and that one has a clear approach to the lexicographic treatment of each and every Zulu word class. Critically analyzing each of these aspects is a massive undertaking, one that cannot be achieved within the ambit of just one research article. The current contribution, therefore, is one in a series.

At face value one would have thought that the logical starting point would have been to discuss macrostructural aspects, and thus to defend the creation of an entire user-friendly lemma-sign list which is word-like rather than stem-like. However, to truly appreciate this effort, it was found that it is more advantageous to analyze the lexicographic treatment of selected Zulu word classes first, and only then to turn to the full macrostructure. As such, De Schryver and Wilkes (2008) concentrated on the treatment of the *possessive pronouns* in a user-friendly Zulu–English dictionary, in this article the focus is on the treatment of *adjectives* in such a dictionary, and in De Schryver (2008a) the focus will be on *quantitative pronouns*.

In order to pick up the thread started in Section 1, and before analyzing the adjectives themselves, the extracts below compare the entries for 'elephant/elephants' in Doke (1)(a) with those in a projected user-friendly Zulu–English dictionary (1)(b).

(1)(a) **-dlovu (indlovu, 2.9.9, izindlovu)** n. [< dlóvu; Ur-B. -yoyú. > umdlovu; indlovudalana; indlovudawana; indlovukazi; indlovunda; indlovundwane.

1. Elephant. *Indlovu iwile, ziphelele zonke izizwe ziye kuxephula kuyo* (The elephant has fallen, and every single one from the tribes has gone to pull off a bit from it; i.e. where the carcass is there will the vultures be gathered together). *Indlovu idla a&asondezeli* (The elephant eats up those who go too near; i.e. don't play with fire). *indlovu enesihlonti* (the elephant with a burning torch — used in *izibongo zikaMbuyaze*).

2. term used of a very stout person.

(1)(b) **indlovu** ** noun 9/10 [^{pl.} **izindlovu**] ► **elephant** ♦ Nansi-ke inganekwane elandwa nguNanana. Kwasukela kwathi indlovu ilambile yahamba ifuna ukudla.

• *Here is the folk tale told by Nanana. Once upon a time, there was a hungry elephant who went looking for food.*

izindlovu plural noun 9/10 See singular **indlovu**

As may be seen from (1)(b), and in contrast to (1)(a), nouns are lemmatized with (and may be found under) their full noun class prefixes, with cross-refer-

ences from the plural to the singular forms.

Extracts (2)(a) and (2)(b) show 'fourth' in the same two dictionaries.

(2)(a) **-ne (isine, 3.2.9, izine)** n. [< adj. **-ne.**]

The fourth place, the fourth. *usuku lwesine* (the fourth day); *ngokwesine* (fourthly).

(2)(b) **isine** * *adverb* ► **fourth** (*used neutrally*) ♦ Uveli uphume isine ku-5000 m. • *Veli came fourth in the 5000 metres.*

♦ [**PC +**]isine ► **fourth** (*used definitely*) ♦ Uginqwe phansi ngomzuliswano wesine. • *He was knocked down in the fourth round.*

The information given under (2)(b) is more explicit — 'spelled out' even — compared to (2)(a). Grammatical guidance is not shunned, and is offered there where the dictionary user will most likely need it (compare this with Sinclair's observation). Here '[PC +]' stands for any prefixed possessive concord. The number of such codes the dictionary user should master has been kept to an absolute minimum.²

A lot more can be said about the lemmatization of the word classes (nouns and adverbs) used as illustrations here, but this will be done in forthcoming studies. Important to note, however, is that all the data shown in (1)(b) and (2)(b) is corpus-driven. The selection of the lemma signs, for instance, is based on overall corpus frequencies, with the top 500 lemmas marked with three stars (***), the next 500 with two stars (**), and the third 500 with one star (*). Meanings have been 'mapped onto use' as seen in the corpus (Hanks 2002). These meanings were then ordered according to individual frequencies and translated into English. Needless to say, the Zulu examples are 'real' (Fox 1987) because they are extracts from the Zulu corpus. For a detailed discussion of the use of this Sinclairian apparatus to dictionary making for the Bantu languages, the reader is referred to De Schryver (2008).

3. True adjective stems in Zulu

Bantu languages have about twenty to thirty so-called 'true adjective stems', and in most existing Bantu dictionaries these are (a) simply (and only) lemmatized as stems, (b) given a basic (or generic) meaning, and, for the larger dictionaries, (c) exemplified with one or more (often invented) phrases. Given Zulu's conjunctive writing system, the required agreement morphemes — known as adjective concords (ACs) — are physically attached to the front of these stems. In such dictionaries, it is thus left to the dictionary user to consult a grammar in addition, where information must be sought on the form and use of the adjective concords, as well as on the morphophonological rules (i.e. sound changes) applicable when attaching an adjective concord to an adjective stem. It is further also assumed that the dictionary user will be able to adapt the meaning depending on class membership of the noun that is being described.

In line with the mission statement presented in Section 2, our claim is that the lemmatization of adjective stems *with their adjective concords* will result in a more user-friendly dictionary. At face value, this may look like a waste of space and resources, as instead of, say, just 25 dictionary articles for adjectives, one will end up with 20 x 25 or thus 500 articles (assuming 16 classes, plus first and second persons). We will come back to this explosion of orthographic forms in Section 4.

At this point, it is instructive to look at the lemmatization of adjectives in a desktop dictionary for Zulu, and to compare the coverage found there with the list of adjectives in a standard Zulu textbook.

Table 1: Adjectives in Zulu: Textbook vs. corpus vs. dictionary data
(with *T* = textbook (Taljaard and Bosch 1993: 99); *Z-E* = user-friendly Zulu-English dictionary; *Freq.* = lemmatized corpus frequency (in 8.5 million words); *Lemma sign*, *POS* and *Translation equivalent(s)* as in Dent and Nyembezi's (1995) dictionary)

T	Z-E	Freq.	Lemma sign	POS	Translation equivalent(s)
✓	✓	5 192	-bi	(adj.)	ugly; bad; evil.
✓	✓	11 667	-bili	(adj)	two.
✓	✓	4 937	-dala	(adj)	old; aged.
-	-	9	-daladala	(adj)	ancient; very old.
✓	✓	4 801	-de	(adj)	long; tall; high; deep.
-	-	4	-fisha	(adj)	short.
-	<i>x-ref</i>	439	-fishane	(adj)	short.
✓	-	18	-fuphi	(adj)	short.
-	-	1	-fusha	(adj)	short.
✓	✓	534	-fushane	(adj)	short.
✓	✓	1 921	-hlanu	(adj)	five.
✓	✓	10 534	-hle	(adj)	good; beautiful; pretty.
✓	✓	18 216	-khulu	(adj)	large; great.
✓	✓	6 875	-ncane	(adj)	small; few; young. <i>kwaba kuncane indawo</i> — keen competition; outcome difficult to predict.
✓	-	69	-nci	(adj)	very small; minute.
✓	-	21	-ncinyane	(adj)	very small.
-	-	3	-ncu	(adj)	minute; very small.
✓	✓	2 818	-ne	(adj)	four.
✓	✓	690	-ngaki	(adj)	how many?
-	<i>der</i>	<i>cf. next</i>	-ningana	(adj)	quite a fair number; not too few.
✓	✓	16 515	-ningi	(adj)	many; much.
-	-	29	-ninginingi	(adj)	numerous.
✓	✓	56 971	-nye	(adj)	some; other.
✓	✓	6 132	-sha	(adj)	new; young.
✓	✓	5 338	-thathu	(adj)	three.
		153 734			

The last three columns of Table 1 list all the adjective stems, 25 in all, as well as their lexicographic treatment, found in the Zulu to English side of Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*. Of these 25 adjective stems, 7 have not been mentioned in Taljaard and Bosch's (1993) *Handbook of isiZulu*, namely the two reduplicated stems *-daladala* (< *-dala*) and *-ninginingi* (< *-ningi*), the

derived stem *-ningana* (< *-ningi*), and the variants *-fusha*, *-fisha*, *-fishane* (~ *-fushane*) and *-ncu* (~ *-nci*). Looking at summed 'lemmatized corpus frequencies', this is defensible, except for *-fishane* and *-ningana*, which are frequent and should have been mentioned. Conversely, these same frequencies also indicate that the forms *-fuphi*, *-nci* and *-ncinyane* are infrequent, so these adjective stems could have been left out as well. This pattern, whereby some frequent forms of a closed set of items are missing while infrequent ones are mentioned instead, is often encountered in textbooks not based on corpus data.³ The first three columns in Table 1 summarize these statistics.

4. Using a corpus to map adjectives onto a user-friendly Zulu dictionary

Given the importance of a corpus within a Sinclairian approach to dictionary making, a few words about the corpus used for this study are necessary. A Zulu corpus totalling 8.5 million running words (tokens) was built, much along the lines described in De Schryver and Gauton (2002: 202-203). A corpus of this size contains a massive 800 000 unique orthographic words (types), of which the top 20 000 were lemmatized. This section represents roughly 70% of the tokens in the Zulu corpus.⁴ Lemmatized corpus frequencies in this article therefore represent the summed frequencies of all items brought together during lemmatization. To complete some of the tables in this article, also lower corpus frequencies are shown (and counted).

In Table 1, one sees that all the lemmatized corpus frequencies together represent about 150 000 running words. Expressed as a percentage of the tokens used for this study, this corresponds to roughly 2.5% of these tokens. Reformulated, this article — which deals with the adjectives in Zulu — is a lexicographic study of about 2.5% of the Zulu lexicon. Conversely, this also means that an average of 2.5 adjectives for each 100 words is used in any spoken or written Zulu.

For the envisaged user-friendly Zulu–English dictionary, the idea is to describe the most frequent 5 000 lemmas only (5 000 in Zulu, and 5 000 in English). The minimum frequency of each Zulu orthographic form before lemmatization was 42, after lemmatization this figure climbed to 50. In other words, the lemmatized corpus frequency must be at least 50 for any Zulu lemma to be considered for inclusion. Applied to the adjectives, one obtains the data shown in Table 2. In this table, the top row lists the various Zulu class numbers as well as the first and second persons singular and plural, 20 in all, while the first column lists the same 25 adjective stems from Table 1.

The ticks (✓) in Table 2 indicate that of the 500 candidate adjectives to be lemmatized, only 160 are left. (Note that the line at *-ningana* was left blank, cf. Section 5.1 below.) Furthermore, given the adjective concords for classes 1 and 3 (and the 2nd person singular) are equal — namely *om(u)-*, as well as those for classes 8 and 10 — namely *eziN-*, and for classes 15 and 17 — namely *oku-*, these 160 collapse to just 126 articles from the point of view of the number of diction-

ary articles. Within a corpus-driven framework, therefore, the explosion of truly important adjectives is not necessarily so dramatic.

Table 2: Adjectival forms in a user-friendly Zulu dictionary with 5 000 lemmas (with *Adj.* = adjective stem; *Cl.* = noun class number and 1st and 2nd persons)

Adj. ↓	Cl. ⇒	1	2	3	4	5	6	7	8	9	10	11	14	15	16	17	18	1sg	1pl	2sg	2pl
-bi		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-bili		-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	✓	-	-	-	-	-
-dala		✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	-	-
-daladala		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-de		✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-
-fisha		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fishane		✓	-	✓	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
-fuphi		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fusha		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-fushane		✓	-	✓	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
-hlanu		-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-hle		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-khulu		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-ncane		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	✓	-	-	-
-nci		-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-
-ncinyane		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-ncu		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-ne		-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-ngaki		-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-
-ningana		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-ningi		✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	-	-	-	-
-ninginingi		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-nye		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-sha		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	-	-
-thathu		-	✓	-	✓	-	✓	-	✓	-	✓	-	-	-	-	-	-	-	-	-	-

5. Advantages of lemmatizing word-like adjectives rather than stems

Observe that 126 entries for adjectives out of a total of 5 000 dictionary articles, corresponds to 2.5% of the total. A word-based approach to the lemmatization of adjectives (in contrast to the traditional stem-based approach) thus also gives a far better reflection of the distribution of the lexicon: Zulu speech and text contains 2.5% adjectives; the number of articles for adjectives in a user-friendly Zulu dictionary is also 2.5%. This finding, of course, is a kind of self-fulfilling prophesy.

There are a number of additional advantages to lemmatizing adjectives with their adjective concords; the main ones are discussed in the next four sections. Each of these sections is accompanied by detailed corpus statistics and star ratings, aimed at shedding further light on the soundness of lemmatizing word-like adjectives. In order not to overload the tables that follow, the ticks (✓) from Tables 1 and 2, which indicated the presence of certain forms, are replaced with the background shading of the corresponding cells (■).

5.1 On varying semantics and diminutives

A first semantic aspect that is lost when one lists adjective stems only, with one overarching meaning (as for instance seen in Table 1), is the different meanings some singular vs. plural forms take.⁵ This is the case for all the adjectives shown in Table 3, and is illustrated for *-nye* in (3).

Table 3: Adjectives expressing 'another/other'; 'much/many'; and 'small/few'

Cl.	AC	-nye	-ningi	-ninginingi	-ncane	-nci	-ncinyane	-ncu
1;3 2sg	om(u)-	omunye 9 103 ***	omningi 168	omninginingi 0;1;0	omncane 1 460 **	omunci 0	omncinyane 3;1	omuncu 0;1;0
2	aba-	abanye 12 908 ***	abaningi 5 267 ***	abaninginingi 1	abancane 735 *	abanci 0	abancinyane 7	abancu 0
4	emi-	eminye 2 211 ***	eminingi 1 477 **	emininginingi 6	emincane 79	eminci 0	emincinyane 1	emincu 0
5	eli-	elinye 4 375 ***	eliningi 46	elininginingi 0	elincane 465 *	elinci 0	elincinyane 1	elincu 0
6	ama-	amanye 4 358 ***	amaningi 1 819 **	amaninginingi 5	amancane 582 *	amanci 0	amancinyane 0	amancu 0
7	esi-	esinye 3 004 ***	esiningi 426	esininginingi 0	esincane 513 *	esinci 0	esincinyane 1	esincu 0
8; 10	eziN-	ezinye 8 418 ***	eziningi 4 513 ***	ezininginingi 10	ezincane 769 *	ezinci 0	ezincinyane 0	ezincu 0
9	eN-	enye 5 569 ***	eningi 627 *	eninginingi 1	encane 1 527 **	enci 0	encinyane 4	encu 2
11	olu-	olunye 322 **	oluningi 82	oluninginingi 0	oluncane 216	olunci 69	oluncinyane 1	oluncu 0
14	obu-	obunye 118	obuningi 47	obuninginingi 0	obuncane 15	obunci 0	obuncinyane 0	obuncu 0
15; 17	oku-	okunye 5 585 ***	okuningi 2 039 **	okuninginingi 1;4	okuncane 427	okunci 0	okuncinyane 1;0	okuncu 0
1sg	engim(u)-	engimunye 0	engimningi 0	engimninginingi 0	engimncane 84	engimunci 0	engimncinyane 1	engimuncu 0
1pl	esiba-	esibanye 0	esibaningi 4	esibaninginingi 0	esibancane 1	esibanci 0	esibancinyane 0	esibancu 0
2pl	eniba-	enibanye 0	enibaningi 0	enibaninginingi 0	enibancane 2	enibanci 0	enibancinyane 0	enibancu 0
	Freq.	56 971	16 515	29	6 875	69	21	3

- (3) **esinye** *** *adjective cl. 7* ► **another (one)** ♦ Isigcino umfana lowo waba yinkosi yesizwe esinye. • *Eventually that boy became the king of another nation.*

ezinye *** *adjective cl. 8, cl. 10* ► **other(s)** ♦ Nazi ezinye izibonelo. • *Here are other examples.* ♦ Zathi ezinye izingane ushaywe nguMdingi. • *Other children said he was hit by Mdingi.*

For classes 15 (the infinitive class) and 17 (the locative class, with 16, 17 and 18 all collapsed into 17) the meaning often deviates even further, as may be seen when comparing (4) with (3).

(4) **okunye** *** *adjective 1 cl. 15* ► **some other; certain** ♦ Izangoma okunye ukudla azikudli, ziyakuzila. • *Diviners do not eat certain foods, they abstain from them.* 2 *cl. 17* ► **something else** ♦ Okunye okubalulekile yindlela umlobi abhala ngayo. • *Something else that is important is the way the writer writes.* ♦ OkukaMagwababa kwaba okunye ngoba yena akadonswanga muntu. • *In Magwababa's case it was something else, because he was not dragged by anyone.*

For *-ncane* one core meaning is present for all classes, but for the plural classes corpus evidence points to an additional meaning. Compare (5) with (6).⁶

(5) **omncane** ** *adjective cl. 1, cl. 3* ► **small; young; little** ♦ Uma unezinyawo ezivuvukele uthete usawoti omncane emanzini uma ugeza. • *If you have swollen feet, pour a little salt in the water when you take a bath.* ♦ Indoda yayinomkhaba omncane. • *The man had a small protruding stomach.*

(6) **abancane** * *adjective cl. 2 1* ► **small; young; little** ♦ Izidakamizwa ziyababulala abadlali abancane. • *Drugs are destroying the young players.* ♦ Uma ungumqeqeshi wabadlali abasebancane kufanele ube nesineke. • *If you are a coach of players who are still young, you should be patient.* 2 ► **a few; a small number** ♦ Siyazi ukuthi unabalandeli abancane. • *We know that she has a few followers.*

Whereas 'another' alternates with 'other' for *-nye*, 'much' alternatives with 'many' for *-ningi*. Recall that Dent and Nyembezi had also listed *-ningana* as an adjective. Actually, this is the diminutive of *-ningi*, and is only frequent enough for classes 8 and 10. Given it is a derivative, it may handily be treated under the form from which it is derived, as shown in (7).

(7) **eziningi** *** *adjective cl. 8, cl. 10* ► **many; a lot of** ♦ Wayevamise ukudla yedwa ezikhathini eziningi. • *He used to eat alone many times.* ♦ Izindlu eziningi azakhekanga kahle emalokishini. • *A lot of houses are not built properly in the townships.* ♦ **eziningana** ► **a small number; quite a few** ♦ Sekuyizikhathi eziningana wena uhlala lapha ekhishini ulinde uZokwenzani. • *It is now quite a few times that you have been sitting here in the kitchen waiting for Zokwenzani.* ♦ Ngenkathi befika egalaji kwakukhona izimoto eziningana. • *When they arrived at the garage there were a small number of cars.*

The frequencies — whether summed or individually — for the adjective stems *-ninginingi*, *-nci*, *-ncinyane* (the diminutive of *-nci*) and *-ncu* clearly indicate that these adjectives should not be entered in a user-friendly dictionary, where one attempts to cover what users are most likely to need. There is one exception, however. Although the frequency of *olunci* is just 3, there are 66 occurrences of this adjective with the associative formative *na-* 'with' prefixed to it. (8), therefore, is a possible treatment.

(8) **olunci** *adjective cl. 11*

- **(lutho) nolunci** ► **small thing** (*always used in negative sentences*) ♦ Akukho lutho nolunci olukhona phakathi kwethu. • *There is not even the smallest thing between us.* ♦ Nya! kungasali nolunci phansi. • *Nothing! Not even the smallest thing must remain on the floor.* ♦ Akukho nolunci olwaluyosindisa uCetshwayo. • *There is absolutely nothing that would have saved Cetshwayo.*

The dictionary article shown in (8) is interesting in various ways. Firstly, note that *olunci* has not been given a meaning — this is in line with its extremely low frequency, combined with the fact that the combination that follows *is* given a meaning. Secondly, every single example in the corpus indicates that the form *nolunci* collocates with *lutho* (< *utho* 'something; anything'), which is either physically present in the sentence or, more often, implied — hence the brackets around *lutho*. Thirdly, *(lutho) nolunci* 'something small' is *only* used in environments with a negative 'semantic prosody' — see Sinclair (1998: 16-22) for the full meaning of this term, and De Schryver (2008: 284-285) for a Bantu-language example. Fourthly, this negative semantic prosody is actually carried over from the noun *utho*, as there is nothing inherently negative about the adjective stem *-nci*. Extra contextual guidance is thus required — achieved by means of the text '*always used in negative sentences*' that follows the translation equivalent. Ample example sentences further illustrate the various ways in which the negativity is brought about — here amongst others by means of a negative copulative (*akukho* 'there is/are no(t)'), a negative verb (*kungasali* 'must not remain') and even a negative ideophone (*nya* 'of nothingness, disappearance, ending, silence').

Clearly, in a stem-based dictionary, where just *-nci* is lemmatized, it is sheer impossible one could have reached this level of customized accuracy. In comparison, (9) reproduces the full entry for *-nci* in the all-encompassing Doke and Vilakazi (1953).

(9) **-nci**, adj. [cf. Ur-B. **-ni** > kancí; ubuncí; ncipha; -ncincincí; -ncinyane.]
Tiny, minute, very small. [cf. *-ncane*, *-ncú*.]

Lemmatizing word-like adjectives, then, allows for far more precise meanings to be conveyed, adapted to the particular class of the adjective. Also, derived adjectives such as diminutives can be described exactly there where they occur.

5.2 On morphophonological rules and augmentatives

Table 4: Adjectives expressing 'big'; 'good' vs. 'bad'; and 'new' vs. 'old'

Cl.	AC	-khulu	-hle	-bi	-sha	-dala	-daladala
1;3; 2sg	om(u)-	omkhulu 3 818 ***	omuhle 1 751 **	omubi 604 *	omusha 1 380 **	omdala 1 423 **	omdaladala 0;4;0
2	aba-	abakhulu 444 *	abahle 264	ababi 144	abasha 817 **	abadala 1 397 **	abadaladala 0
4	emi-	emikhulu 316	emihle 410	emibi 250	emisha 165	emidala 48	emidaladala 1
5	eli-	elikhulu 2 198 ***	elihle 1 382 **	elibi 403	elisha 576 *	elidala 249	elidaladala 3
6	ama-	amakhulu 875 **	amahle 480 *	amabi 166	amasha 571 *	amadala 243	amadadaladala 0
7	esi-	esikhulu 1 658 **	esihle 1 206 **	esibi 447 *	esisha 281	esidala 149	esidaladala 1
8; 10	eziN-	ezinkulu 1 162 **	ezinhle 1 096 **	ezimbi 469 *	ezintsha 563 *	ezindala 257	ezindaladala 0
9	eN-	enkulu 4 213 ***	enhle 2 081 **	embi 740 *	entsha 1 086 **	endala 702 *	endaladala 0
11	olu-	olukhulu 1 225 **	oluhle 184	olubi 372	olusha 350	oludala 180	oludaladala 0
14	obu-	obukhulu 277	obuhle 240	obubi 106	obusha 93	obudala 66	obudaladala 0
15; 17	oku-	okukhulu 2 026 **	okuhle 1 438 **	okubi 1 487 **	okusha 250	okudala 63	okudaladala 0
1sg	engim(u)-	engimkhulu 1	engimuhle 1	engimubi 1	engimusha 0	engimdala 149	engimdaladala 0
1pl	esiba-	esibakhulu 2	esibahle 1	esibabi 2	esibasha 0	esibadala 8	esibadaladala 0
2pl	eniba-	enibakhulu 1	enibahle 0	enibabi 1	enibasha 0	enibadala 3	enibadaladala 0
	Freq.	18 216	10 534	5 192	6 132	4 937	9

The data in Table 4 was presented first to see whether or not readers would notice that the form of the stem *-khulu* 'big; large; great' for classes 8 to 10 has changed. Indeed, one of the morphophonological rules in Zulu forbids the succession of **n + kh**, with the result that the **h** is dropped. Likewise, **n + sh** is not allowed, so a **t** is inserted between the *N* of the adjective concord and the initial consonant of the adjective stem. This affects *-sha* 'new; young' for classes 8 to 10. Rather than expecting that dictionary users remember such rules, lemmatizing word-like adjectives immediately gives them the correct forms, as seen in (10) and (11).

- (10) **enkulu** *** *adjective cl. 9* ► **big; large; great** ♦ Ubaba wenze ibhena enkulu yesikole. • *Father made a big banner for the school.*
 ♦ **enkulukazi** ► **very big; very large; very great; huge** ♦ Babulale inyoka enkulukazi, ngiyabona yinhlwathi. • *They killed a very large snake; I think it is a python.*

- (11) **ezintsha** * *adjective cl. 8, cl. 10* ► **new; young** ♦ Kuzokwakhiwa izibhedlela ezimbili ezintsha eSoweto. • *Two new hospitals will be built in Soweto.* ♦ Batsheleke izimali emabhange ukuze bathenge lezi zimoto ezintsha. • *They borrowed money from the banks in order to buy these new cars.*

Other 'orthographic rules' which were implicit so far concern *N* for classes 8 to 10 — *m* before **b** or **f**, *n* elsewhere; and the form of the adjective concord for classes 1 and 3 (and the 2nd person singular) — *omu-* vs. *om-*, as well as the form for the 1st person singular — *engimu-* vs. *engim-*. The first prefix in each series is used for monosyllabic stems, the second for polysyllabic stems. See for instance (12) and (13), respectively (14) and (15), applied to the adjective stems *-hle* 'good; beautiful; nice' and *-bi* 'bad; ugly; evil' vs. *-dala* 'old'.

- (12) **omuhle** ** *adjective cl. 1, cl. 3* ► **good; beautiful; nice** ♦ Ngitshele konke mama wami omuhle. • *Tell me everything my good mother.* ♦ UKhanyi unomzimba omuhle. • *Khanyi has a beautiful body.*
- (13) **omubi** * *adjective cl. 1, cl. 3* ► **bad; ugly; evil** ♦ Nguye umuntu omubi kubalandeli baleli qembu. • *He is the bad person among the followers of this group.* ♦ Uwuthandelani umdlalo omubi kangaka? • *Why do you like such an ugly game?*
- (14) **omdala** ** *adjective cl. 1, cl. 3* ► **old** ♦ Umuntu omdala akanawo amandla okuphikisana nomuntu osemusha. • *An old person doesn't have the strength to compete with a person who is still young.* ♦ Ubona lapho umgwaqo omdala. • *She sees an old road there.*
- (15) **engimdala** *adjective 1p sg* ► **I who am old** ♦ Wacabanga ukuthi ngimdala. • *She thought that I am old.* ♦ Wayesethi: 'Bheka, sengimdala, angilwazi usuku lokufa kwami.' • *And then he said: 'Look, although I am already old, I do not know the day of my death.'*

Note that adjectives for the first and second persons singular and plural are *very rare* overall. There are just 9 in all for the 1st person singular, 18 for the 1st person plural, and 8 for the 2nd person plural.⁷ Finding 2nd person singular adjectives is very difficult, given that the orthographic form of these is the same as for class 1 and 3 adjectives. They probably have the same order of magnitude as the other first and second person adjectives.

As was the case for the reduplicated stem *-ninginingi* 'numerous' (< *-ningi* 'much/many'), also the frequency of the reduplicated stem *-daladala* 'ancient' (< *-dala* 'old') is too low for it to be included in a dictionary covering the most frequent words only.

Further note that (10) above also listed *enkulukazi* 'very big; very large; very great; huge' as a derivation. Indeed, with adjectives the suffix *-kazi* is used for augmentative purposes. Augmentative adjectives being rather rare (cf. Gau-

ton, De Schryver and Mohlala 2004: 374), they can again best be included directly under those adjectives with which they actually occur.⁸

5.3 On class restrictions

The next group of adjectives is peculiar because they only occur with certain classes, namely the plural classes 2, 4, 6, 8 and 10, as seen in Table 5.⁹

Table 5: Adjectives expressing 'how many?', 'two', 'three', 'four', and 'five'

Cl.	AC	-ngaki	-bili	-thathu	-ne	-hlanu
1;3;2sg	om(u)-	—	—	—	—	—
2	aba-	abangaki 199	ababili 1 973 **	abathathu 727 *	abane 382	abahlanu 256
4	emi-	emingaki 124	emibili 1 705 **	emithathu 993 **	emine 542 *	emihlanu 677 *
5	eli-	—	—	—	—	—
6	ama-	amangaki 112	amabili 4 276 ***	amathathu 1 856 **	amane 1 034 **	amahlanu 575 *
7	esi-	—	—	—	—	—
8;10	eziN-	ezingaki 252	ezimbili/mbili 3 380 ***/120	ezintathu 1 725 **	ezine 851 **	ezinhlanu 407
9	eN-	—	—	—	—	—
11	olu-	olungaki 0	olubili 0	oluthathu 1	olune 0	oluhlanu 0
14	obu-	obungaki 1	obubili 7	obuthathu 8	obune 0	obuhlanu 0
15;17	oku-	okungaki 1;1	okubili 0;86	okuthathu 1;27	okune 2;7	okuhlanu 0;6
1sg	engim(u)-	—	—	—	—	—
1pl	esiba-	—	—	—	—	—
2pl	eniba-	—	—	—	—	—
	Freq.	690	11 667	5 338	2 818	1 921

Clearly, one cannot 'count' singular things, so the distribution seen in Table 5 is not so surprising. This said, when assigning a meaning to adjective stems in isolation, without truly considering all and only those *possible* forms that belong to the paradigm, it is rather easy to err in this regard. Taljaard and Bosch (1993: 99), for instance, assign the meaning 'how much/many?' to *-ngaki*. This is incorrect, as 'how *much?' would only be used for singular adjectives, of which there are none for this adjective stem! Compare with the adjective stem *-ningi* 'much/many' in Section 5.1 which, conversely, does have both singular and plural forms. (16) shows a possible treatment for one of the forms of *-ngaki* 'how many?'

- (16) **emingaki** *adjective cl. 4* ► **how many?** ♦ Linemibala emingaki ifulegi laseNingizimu Afrika? • *How many colours does the South African flag have?*
 ▪ **iminyaka emingaki** ► **how old?** ♦ Waqala uneminyaka emingaki ukucula?
 • *How old were you when you started singing?*

In (16) one can also see how frequent combinations may be included in a user-friendly dictionary — again directly under the relevant lemma (here: 'how old?' < 'how many years?', with 'years' a plural noun in class 4).

The other forms in Table 5 are used for counting: *-bili* 'two', *-thathu* 'three', *-ne* 'four' and *-hlanu* 'five'. An extra morphophonological rule applies here: in the combination **n** + **th**, the **h** needs to be dropped. This affects *-thathu* in classes 8 and 10. Interestingly, going from 2 to 5, the overall frequency decreases. People seem to talk more often about a few things rather than about many things. (17) is a straightforward example.

- (17) **ezimbili** *** *adjective cl. 8, cl. 10* ► **two** ♦ Kwakuxabana ziphi lezi zizwe ezimbili kule mpi? • *Which of these two countries were at loggerheads in this war?* ♦ Izimbuzi ezimbili ngezami. • *The two goats are mine.*
mbili *adjective cl. 8, cl. 10* Short form of **ezimbili**

Corpus evidence indicates that *mbili* (frequency = 120), a short form of *ezimbili*, is frequent enough to be lemmatized. A straightforward cross-reference to the full form suffices here, see (17). Needless to say, a form such as *mbili* is not lemmatized nor covered in traditional Zulu dictionaries.

Only one adjectival form 'breaks' the symmetrical pattern seen in Table 5, namely *okubili*, for class 17, see (18).

- (18) **okubili** *adjective cl. 17* ► **two things** ♦ Benzani laba bafana? Shono okubili. • *What are these boys doing? Name two things.* ♦ Nokho kubili athanda ukukugqamisa lapha. • *Nevertheless, there are two things that he wants to highlight here.*

Two reasons may be offered for the relatively high frequency of *okubili*, the first being that people tend to count up to two rather than higher, the second being that this effect is doubled as a result of the copulative use (cf. Section 6 below).¹⁰

5.4 On cross-references

The adjectives *-fushane*, *-fishane*, *-fusha*, *-fisha* and *-fuphi* may all be used to refer to 'short' people or things. The last three, however, are clearly not frequent enough to be included in even the larger Zulu dictionaries. The first two are synonyms of one another, and overall summed frequencies indicate that *-fishane* should be considered a variant of *-fushane*.¹¹ All this information, then, leads to a treatment like (19).

- (19) **esifishane** *adjective cl. 7* = **esifushane**
esifushane *adjective cl. 7* ► **short; brief** ♦ Isitimela sithatha isikhathi esifushane uma sisuka eDanawozi bese sifika eGlencoe. • *The train takes a short time going from Dannhauser to Glencoe.*

So far, the following 'opposite adjective pairs' were discussed: *-khulu* 'big' and *-ningi* 'much/many' vs. *-ncane* 'small/few'; *-hle* 'good' vs. *-bi* 'bad'; and *-sha* 'new' vs. *-dala* 'old'. As the last in this series, *-de* 'long' may be contrasted with *-fushane* 'short'. See Table 6 for the full picture, and (20) for one example.

Table 6: Adjectives expressing 'long' vs. 'short'

Cl.	AC	-de	-fushane	-fishane	-fusha	-fisha	-fuphi
1;3;2sg	om(u)-	omude 304	omfushane 60	omfishane 65	omfusha 0	omfisha 1;1;0	omfuphi 1;2;0
2	aba-	abade 39	abafushane 6	abafishane 3	abafusha 0	abafisha 0	abafuphi 1
4	emi-	emide 111	emifushane 30	emifishane 37	emifusha 0	emifisha 0	emifuphi 0
5	eli-	elide 362	elifushane 20	elifishane 12	elifusha 0	elifisha 0	elifuphi 5
6	ama-	amade 210	amafushane 19	amafishane 29	amafusha 0	amafisha 0	amafuphi 3
7	esi-	eside 2 571 ***	esifushane 98	esifishane 56	esifusha 1	esifisha 0	esifuphi 0
8;10	eziN-	ezinde 237	ezimfushane 102	ezimfishane 63	ezimfusha 0	ezimfisha 0	ezimfuphi 0;1
9	eN-	ende 637 *	emfushane 189	emfishane 160	emfusha 0	emfisha 1	emfuphi 3
11	olu-	olude 241	olufushane 7	olufishane 9	olufusha 0	olufisha 0	olufuphi 1
14	obu-	obude 54	obufushane 1	obufishane 1	obufusha 0	obufisha 0	obufuphi 0
15;17	oku-	okude 34	okufushane 1;0	okufishane 1;3	okufusha 0	okufisha 1;0	okufuphi 1;0
1sg	engim(u)-	engimude 0	engimfushane 1	engimfishane 0	engimfusha 0	engimfisha 0	engimfuphi 0
1pl	esiba-	esibade 0	esibafushane 0	esibafishane 0	esibafusha 0	esibafisha 0	esibafuphi 0
2pl	eniba-	enibade 1	enibafushane 0	enibafishane 0	enibafusha 0	enibafisha 0	enibafuphi 0
	Freq.	4 801	534	439	1	4	18

(20) **amade** *adjective cl. 6* ► **long; tall; high; deep** ♦ Abantu abaningi basebenza amahora amade kodwa bahola amakinati. • *Many people work long hours but earn peanuts.*

6. Qualificative adjectives versus copulative adjectives

In the picture sketched so far, although dealing with complex issues already, a few extra parameters have purposely been avoided. Firstly, in all but three of the examples from (3) to (20), the orthographic form illustrated in the example sentences is exactly the lemma sign. As a result, it may now appear as if the lemma signs are also the only members of each paradigm. Of course, this is not the case.

During dictionary compilation, the lexicographers have at their disposal the full list of all the forms which were brought together during lemmatization, as well as the frequencies for each of these forms. For instance, for *abancane*, see (6) above, these forms are:

(21) abancane <483>, abasebancane <135>, nabancane <66>, besebancane <51>

As one can see, here the most frequent form of the lemma (*abancane*, with a frequency of 483) equals the lemma sign (*abancane*, with a summed lemmatized frequency of 735). This pattern is seen for 113 of the 126 adjectives. In other words, for about 90% of the adjectives, the lemma sign is also the most frequent form of the adjective. This, then, is another good and user-friendly consequence of lemmatizing adjective stems with their full adjective concords.

Rather than choosing random forms to illustrate the lemma signs, the lexicographers try to pick frequent forms from lists such as (21). If one now returns to the article shown in (6), then one notices that the second example exemplifies the second-most frequent form of the lemma, namely *abasebancane*. This form can be analyzed as follows: *aba-* (relative concord class 2, RC2) + *se-* (progressive formative) + *ba-* (adjective prefix class 2, AP2) + *-ncane* (adjective stem) 'who are still small/young/little'. Hence the example: *Uma ungumqeqeshi wabaddali abasebancane kufanele ube nesineke*. 'If you are a coach of players who are still young, you should be patient.'

The last form in (21), *besebancane*, is actually a *copulative adjective*. This is the second aspect that has been kept out of the discussion so far. Under 'adjectives', then, both the qualificative (i.e. the form with the adjective concord) and the copulative uses are brought together. In some rare cases, a copulative adjective is even more frequent than its corresponding qualificative adjective. In (22), for instance, the frequencies are: *bahle* <153>, *abahle* <111>; which explains the order of the examples.

(22) **abahle** *adjective cl. 2* ► **good; beautiful; nice** ♦ Abantu besifazane bahle ngezindlela ezingafani. • *Women are beautiful in different ways.* ♦ Khetha abangani abahle abangeke bakudukise. • *Choose good friends who will never lead you astray.*

In (18) above, the lemma was formed from: *okubili* <49>, *kubili* <37>; which should again make the treatment clear.

Also above, (15) is an extreme case. The forms brought together during lemmatization are: *ngimdala* <73>, *sengimdala* <72>, *engimdala* <4>. In other words, the qualificative entry was 'created' to cater for the two copulative uses. Both examples, of course, only illustrate the copulative uses (with 'se' in *sengimdala* the auxiliary verb *-se*).

To all intents and purposes both qualificative and copulative adjectives may be covered by the same translation equivalents (even though the copulative use includes the meaning 'to be' in addition). To turn a qualificative adject-

tive into a copulative adjective it suffices to drop the initial vowel for all classes, except for class 9 where the initial **e** becomes an **i** (cf. Section 7). This is a feature that *can* and *must* be explained in the integrated 'corpus-based dictionary mini-grammar' (compare with De Schryver and Taljard 2007). It must be explained, because a user who encounters a copulative use of an adjective will need to be able to add the initial vowel in order to look up the lemmatized qualificative use.

7. The tension between linguistics and lexicography

It is now time to depart from the gentle linguistic introduction which has characterized the discussion so far, and to look at some hardcore linguistic facts. What is really the case with the adjective in Zulu? One first needs to know that the **adjective concord** (AC) is actually composed of two formatives, the **relative concord** (RC) plus the **adjective prefix** (AP):

$$(23) \quad AC = RC + AP$$

The RC is the *abbreviated* RC. The RC itself is formed by prefixing the relative formative **a-** to the **subject concord** (SC). As such, one for instance obtains *aba-* for class 2 (< *a-* + *ba-*, abbreviated form: *a-*), or *e-* for class 9 (< *a-* + *i-*). The AP for class 2 is *ba-*, so the AC for this class becomes *aba-* (< *a-* + *ba-*); the AP for class 9 is *iN-*, so the AC for this class becomes *eN-* (< *e-* + *iN-*).

With 'AStem' the **adjective stem**, the basic structure of a **qualificative adjective**, respectively **copulative adjective** is:

$$(24) \quad \begin{aligned} \text{Basic qualificative adjective} &= AC + \text{AStem} \\ \text{Basic copulative adjective} &= AP + \text{AStem} \end{aligned}$$

In other words, to turn a qualificative adjective into a copulative adjective, one basically drops the RC. For instance, in (22), *abahle* 'good' becomes *bahle* '(they) are good'. Likewise, the form from (10), *enkulu* 'big', becomes *inkulu* '(it) is big'.

This brief sketch summarizes most adjectival forms seen so far. These forms can however also be preceded by various other prefixes. In order to streamline the presentation, we can divide these into three groups. Firstly, qualificative adjectives may be preceded by a **possessive concord** (PC):

$$(25) \quad PC + AC + \text{AStem}$$

Secondly, the qualificative adjectives can also be preceded by any of the following formatives: **locative** (*kwa-/ku-*), **associative** (*na-*), **instrumental** (*nga-*), **comparatives** (*kuna-* (< *ku-* + *na-*), *njenga-*), and combinations thereof (attested for the top adjectives are: *ngakwa-/ngaku-* (< *nga-* + *ku-*), *nakwa-* (< *na-* + *ku-*), *nanga-* (< *na-* + *nga-*)).

For instance, all the forms seen at the bottom of (26) are also the forms seen by the lexicographers during dictionary compilation in *TshwaneLex* (for more on this software, cf. Joffe et al. 2008). An analysis is shown in (27).

- (26) **abanye** *** *adjective cl. 2* ► **other(s)** ♦ Akufanele uhleke abanye abantu uma bengaphumeleli empilweni. • *You should not laugh at other people who are not succeeding in life.*

12908, abanye <8480>, nabanye <2243>, kwabanye <977>, ngabanye <548>, yabanye <158>, zabanye <149>, kunabanye <136>, njengabanye <111>, nakwabanye <106>

- (27)
- | | | |
|-------------|---------------------------|-----------------------------------|
| abanye | = AC2 + AStem | = 'other(s)' |
| nabanye | = ass + AC2 + AStem | = 'and/with other(s)' |
| kwabanye | = loc + AC2 + AStem | = 'to other(s)' |
| ngabanye | = instr + AC2 + AStem | = 'concerning/with other(s); ...' |
| yabanye | = PC4or9 + AC2 + AStem | = 'of other(s)' |
| zabanye | = PC8or10 + AC2 + AStem | = 'of other(s)' |
| kunabanye | = comp + AC2 + AStem | = 'than other(s)' |
| njengabanye | = comp + AC2 + AStem | = 'just like other(s)' |
| nakwabanye | = ass + loc + AC2 + AStem | = 'and to other(s)' |

Thirdly, corpus evidence — as summarized in the bottom slots such as the one seen in (26) — further indicates that all the structures shown in (28) are possible (this is a selection of ten only).

- (28)
- RC + progressive formative *se-* + AP + AStem
 - RC + negative formative *nge-* + AP + AStem
 - RC/SC + copulative formative *ng-* + AC + AStem
 - SC in situative mood + progressive formative *se-* + AP + AStem
 - (SC in situative mood + auxiliary verb *-se* +) SC in situative mood + AP + AStem
 - SC in remote past tense (+ auxiliary verb *-se*, optionally dropped) + SC in situative mood + AP + AStem
 - SC + potential formative *nga-* + AP + AStem
 - negative morpheme in indicative mood *a-* + SC in indicative mood + AP + AStem
 - auxiliary verb *-se* (+ SC, obligatory in situative mood) + AP + AStem
 - [for class 9] copulative formative (terminal depressor) *y-* + AP9 *iN-* + AStem

As a random example, some of the forms from (29) are analyzed in (30).

- (29) **omunye** *** *adjective cl. 1, cl. 3* ► **another (one)** ♦ Kubulawe omunye osomatekisi KwaZulu-Natali. • *Another taxi man was murdered in KwaZulu-Natal.* ♦ Umehluko omunye ukuthi wayengasakhulumeli futhi. • *Another difference is that she was no longer talkative.*

9103, omunye <5608>, komunye <906>, ngomunye <729>, nomunye <519>, ungomunye <515>, ongomunye <215>, ngingomunye <89>, lomunye <85>, yomunye <79>, njengomunye <73>, engomunye <70>, wayengomunye <66>, ubengomunye <56>, womunye <47>, kunomunye <46>

- (30) ungomunye = SC1or3 *u-* + copulative formative *ng-* + AC1or3 *omu-* + AStem *-nye*
 = 'he/she/it is one of the others'
 ongomunye = RC1or3 *o-* + copulative formative *ng-* + AC1or3 *omu-* + AStem *-nye*
 = 'he/she/it is another one'
 wayengomunye = SC1 in remote past tense *wa-* + time auxiliary *-be* (dropped here) + *-y-* (bridging sound) + SC1 in situative mood *e-* + copulative formative *ng-* + AC1 *omu-* + AStem *-nye* = 'he/she was another one' (in the remote past tense)
 ubengomunye = SC1 in present tense *u-* + time auxiliary *-be* + SC1 in situative mood *e-* + copulative formative *ng-* + AC1 *omu-* + AStem *-nye* = 'he/she was another one' (in the near past tense)

When considering the examples listed under (27) and (30) — which, it must not be forgotten, are but a tiny selection of the full spectrum —, it is easy to understand why traditional lexicographers for the Bantu languages in general, and for Zulu in particular, decided to collapse all of these forms into just 'a single dictionary article', here the adjective stem *-nye*. Lemmatizing *all* forms, even only all *frequent* forms, remains an impossibility. Yet, meeting the dictionary user halfway is a realistic proposition, as has been shown in Section 5.

Indeed, as is obvious from the full statistics listed under (26) and (29), lemmatizing the basic qualificative adjectives only, truly covers the most important uses. The other forms may be relegated to the integrated mini-grammar. There too, corpus statistics with regard to the frequency of the various *structures* may be used in the endeavour to present the core issues.

The tension, then, between a detailed, all-encompassing linguistic coverage on the one hand, and a user-friendly, tailored lexicographic treatment on the other, has been eased by a study of overall corpus statistics. What is of prime importance ends up in the dictionary A-to-Z section; what is secondary ends up in the attached grammar.

8. Getting the adjective frequencies right

Frequencies such as those shown in the two previous sections are not always as straightforward as they may seem. At face value, several adjectival forms may also be other parts of speech. When one actually sets out to compile a dictionary article, it is not exceptional to browse through literally hundreds of concordance lines in order to extract the meaning(s) and to select appropriate example sentences for the lemma one is working on. However, when one needs to get an idea of the relative frequencies of different forms — be these on homonym level, sense level, or both simultaneously —, sampling techniques are used for all frequent items in order to limit the number of concordance lines to be studied.¹² Typically, the lexicographers aim at studying about fifty KWIC lines at this point. In Figure 1, *okudala* is being analyzed, an item which can be both an adjective and a verb (marked with 'a' and 'v' respectively during the analysis).

N	Concordance	Set	File
13	e-UK uMax Jones, uthu lokhu okwenzekile kube wukuvuka kokulimala okudala osekumhlophe isikhathi eside lo mgijimi. "Usebe nale nkinga i	a	is200407.txt
14	Izinkomba zivamise ukubangwa ukumila noma ukuvuvuka komthambo okudala ukuthikameza kokuhamba komchamo. Izinkomba zomdluza	a	lsdbzulhe.txt
15	si kwaso isithombe lesa kubhalwe nje ukuthi: "Kwabanye base kuvuke okudala! Yeka lezozinsuku ezadlulayo eMpumelelo!" Noma engabhal	a	ingiyeken.txt
16	xoye ezinye. Izinto zomhlaba ziyayaka. Ukwenza komhlaba kuyayaka. Okudala kwawo akwejoye. Okusha kwawo kuyadida. Kungaba kh	a	manqamp.txt
17	u lwakho lokuzalwa namuhla? Sekumele uphothule ngemfanelo lokho okudala ngaphambi kokuqalisa okusha ezinhlweni zakho. Abanye ox	a	is200410.txt
18	zaga; futshi zona zihlezi ziphenduphenduka nesikhathi. Yikho-ke lokhu okudala ukuthi kabekhona izisho ezintsha; zibe zingekho izaga ezintsh	v	1izinhlan.txt
19	iphenduka ibe ngumfana wami esiphicaphicwani. Yikho kanye lokhu okudala ukuba kube khona ukucashelana okonophelo eziphicaphic	v	langigqib.txt
20	ba evuka izimbongi zohlanga ezibongela amakhosi. Yikho kanye lokhu okudala ukuba ukuhaywa kwezithakazelo kube sengathi umndeni usuk	v	1izinhlan.txt
21	ubha, ethi kuhle ngivume uma lento yenziwe yimi. Ngaphika ngahlanza okudala. Ayibize nenkanyamba abuze ukuthi yona icabangela kubani.	a	1Inasi-kei.txt
22	kwakuyindoda elungile neyayinomqondo. Ngahlala naye ngakumbula okudala. Kwakummandi silezi emthunzini womsimbithi sinatha siphun	a	ingisinga.txt
23	gise amaphutha esipelingi. (c) Ukuphimsa amagama ngakungafanele okudala ukubhaleka kwegama kabi. 21 (d) Ukuswela ulwazi olunzulu	v	1isizulu9.txt
24	be ubuza ngoba uthuywe ngabaseshi-ke, mina ngiyophika ngihlanza okudala qiniso. DOLLY: Ungibona nje sengingaba yimpimpi yabases	v	mavenge.txt
25	celemba begawula namazwi abo elapha phezulu eshikisha ngokudala. Okudala lokhu kwakuyini? Angazi. Okudala kwakusho ukucima kwama	a	inje-nemp.txt
26	a ukwelapha isifo esithize? Wukungazi kuhambisana nokukhukhumala okudala umqondo orjena kubantu bakithi. Inkungu ibhokile. Sikhukhu	a	is200602.txt
27	sebona ngokocutha kwamadebe nje. Zasukelana futshi. Kwathi nintini okudala kwabona ukuthi akusizi, akuzame njengakuqala. Nempela shi,	a	emhlabe2.txt
28	miny e imidlalo lesisikhathi siphawuleka kalula ngoba kwenzeka okuthile okudala udweshu. Nokho kweminye imidlalo akubulula ukwazi ukuthi les	v	1isizulu7.txt
29	Ayengezwani neze naleligama amaZulu uma sekukhona okuwadidayo okudala ukufa, kube kunyama ngoba kuyisebusuku. Abantu bakhe u	v	hudeman.txt
30	mkhulu nokhoko nokhuluhlwanane nadalwa nguNkulunkulu onamandla okudala, okuphilisa nokubulala. Yingakho badalwa nje. baphila, bafa,	v	ingevuye.txt
31	mbe esisekhasini eIngaphandle sikujabulise lapho kuthiwa sekuyeka okudala sebekhumbula lezonsuku ezadlulayo eMpumelelo. kodwa b	a	ingiyeken.txt
32	amba ngale ndlela? Lesi senzo sicekela phansi isithunzi senhlangano okudala ukuthi amalungu aphelelwe wumdlanla nentshisekelo yokuzi	v	is200505.txt
33	lwa kulona elinezinkaba zethu. Sike sibheke emumva uma sikhumbule okudala, sikubone kufana nomhwaluko. Kuthi uma sisinga phambili	a	ingisinga.txt
34	odwa elidala. ipaki elihle, uthi olutshane, ubahlalu obumblophe, ukufa okudala. 1 kuyanda manje 2 sithela kancane 3 buthengwa lapha. 4 li	a	1indlela2.txt
35	nayo. Qala lapho sekonakele khona izinto, sekukhona ukungubuzana okudala abalingiswa babhekane nezinkinga eziningi. Izigcawu: Isigca	v	lubhagatw.txt
36	waqedelela i-free kick eshaye ipali yabuyela enkundleni. Ukuzimisela okudala kukaMokoena ayekuhombisa kuMaGlug-Glug akusabonakali	a	is200511.txt
37	ele. Ukusungula nokucabanga okusha akusho ukuthi sekumele ulibale okudala. Taurus: Apr 21 - May 21 Kuzomele ukhumbule ukuthi ukual	a	is200412.txt
38	buzi, ndodana! Ngabe ngubani lowo? Sipikili: Kukhona omunye umlisa okudala sisebenza naye laphayana kumkhwenya wethu. Nami ngafika	a	1inkundla.txt
39	ogani wakhe ngaphambi kokuba ale indima: "Yebo-ke MaSibisi, uvuse okudala namhlanje mngani wami. Awungitshale ukuthi uvuke nini ngob	a	lakuyiwe.txt
40	obe lokungibulala neqembu lakho." Uthi waphika uShamase wahlanza okudala wathi: "Ndabezitha angikwazi lokho." Uthi yamfutha iNkosi yat	a	nyambos.txt
41	bane; nalenombazana efuleyo angiyazi." Waphika uKhanyile wahlanza okudala. Wahleka umseshi wanikina ikhanda, waseguqukela kuSigab	a	ubogawul.txt
42	ajabula wafa uMcineleli. Kwathi lapho esethi uyayihlala yaphika wahlanza okudala intombi. Kulapho-ke kwaboboka khona ithumba. "Cha, ang	a	ubogawul.txt
43	wenzeka uma umzimba unokuphikisana. Amehlo nobuchopho yikhona okudala lokhu kakhulu. Amehlo asuke ekhomba ukuthi umzimba uyah	v	is200510.txt
44	i wemoto kanye nabalapho esuke iyokwenziswa khona iservice yikhona okudala ukuqala kwezinkinga. UVan Zyl uthi zonke izimoto ezintsha zi	v	is200511.txt
45	ela, aqome khona ukuba afe kunokuba alahlakelwe yileyo. Yilokho okudala usizi nomunye ngoba usuke engenyana umuntu. omubi futshi us	v	1isizulu7.txt
46	uhlanganisile nje amakhubalo ezinhlobonhlobo, nezikhumba zezilwane okudala zafa. Kanti futshi nasemsamo laphaya kuthule nje kuthe du, am	a	mcebobo.txt
47	hi kanti sengiguge ngempela." "Kanti-ke lutho kawugugile wena Zondi. Okudala ukuba sikhohlane ukuthathwa yimisebenzi kanye nokwahluka	v	lamahlaya.txt

Figure 1: Sampling *okudala*, which is both an adjective (a) and a verb (v)

In Figure 1, the corpus software, *WordSmith Tools* (Scott 2008), was requested to randomly select one out of every three occurrences, and the allocation seen in the sample was then used to distribute the total frequency across the verb *-dala* 'create', and the adjective *okudala*, shown in (31).

- (31) **okudala** *adjective* 1 cl. 15 ► **old** ♦ Ukuzimisela okudala kukaMokoena ayekuhombisa kuMaGlug-Glug akusabonakali njengoba emaningi amaphutha awenzayo. • *The old determination of Mokwena which he had shown with the Team of the Crocks is no longer visible because of the many mistakes that he made.* 2 cl. 17 ► **something old; long ago** ♦ UNondela wayesekhumbule okudala ngempela kusabusa inkosi uNdaba. • *Nondela had remembered the really old things during the reign of chief Ndaba.* ♦ Kukhona omunye umlisa okudala sisebenza naye laphayana. • *There is another male person with whom we worked together long ago.*

Focusing on the adjective: The meanings for the different senses were 'derived' from the corpus, and at the same time one of course keeps an eye on all other items within the same paradigm too — compare for instance (14) and (15). Further observe that two of the three examples in (31) were also selected from the sample seen in Figure 1 (viz. lines 36 and 38). As another example, the frequency of *kubili*, see (18), was split over the adjectival and nominal use.

9. Pinpointing idiomatic uses with adjectives

In Table 1, one could see that Dent and Nyembezi (1995) covered one instance of idiomatic use with an adjective, reprinted in (32).

- (32) **-ncane** (adj) small; few; young.
kwaba kuncane indawo — keen competition; outcome difficult to predict.

Coverage of idiomatic use is of course commendable, but in a user-friendly dictionary, this usage should at least be truly frequent too. A corpus-wide search through 8.5 million words of Zulu returns just six instances of *-ba/-be kuncane indawo*. The meaning 'keen competition' cannot be derived from these lines, however, rather something like 'it is not comprehensible what the outcome will be'. The latter is also the meaning listed in Nyembezi's (1992: 317) monolingual dictionary *Isichazimazwi sanamuhla nangomuso*, as well as in Nyembezi and Nxumalo's (1966: 223) miscellany of Zulu culture *Inqolobane yesizwe*. In any case, there are certainly better candidates; (33) is an example.

- (33) **oludala** *adjective cl. 11* ► **old** ♦ Indibilishi nosheleni uhlobo oludala lwemali. • *A penny and a shilling are an old type of money.*
 ▪ **kusadliwa ngoludala** ► **old customs are still followed** (*Literally: there (things are) still being eaten with an old one (referring to a spoon)*) ♦ Kusadliwa ngoludala eMsinga. • *Old customs are still followed at Msinga.*

While the frequency of *oludala* is 60, that of *ngoludala* is twice as high, 120. Of these 120 all but one of the occurrences refer directly to the idiomatic use. The adjective *oludala*, then, has clear open and idiomatic uses, roughly one-third being open, two-thirds being idiomatic (compare with Sinclair 1987: 319-320).

10. Overruling strict principles for the sake of user-friendliness

A lexicographer's job is one of repetitious systematicity. Every now and then, however, flexibility is called for in the user's interest. (34) is a case in point.

- (34) **okuhle** ** *adjective* Compare **kuhle**¹ *cl. 15* ► **good; beautiful; nice** ♦ Bamfisela ukuhlolwa okuhle. • *They wished him a good examination.* *2 cl. 17* ► **something good / beautiful / nice** ♦ Siyifisela okuhle le ngane. • *We wish this child good luck.*
 ▪ **okuhle kodwa** ► **only the best** ♦ Umfisele okuhle kodwa nempilo ende. • *He wished her only the best and a long life.*
 ♦ **kungakuhle** ► **it would be good / beautiful / nice** ♦ Kungakuhle uma uthisha engabanika amaphuzu aphezulu. • *It would be nice if the teacher could give them high marks.*
Note: For the copulative use of 'okuhle', see 'kuhle' (it's good/beautiful/nice).

1438, okuhle <964>, kungakuhle <253>, kukuhle <175>, ngokuhle <46>

The article shown in (34) has a bit of everything: two senses (one for class 15, one for 17), a frequent collocation (*okuhle kodwa*), and a frequent derivation (*kungakuhle*). What this article does not cover is the copulative use. In this one exceptional case, the copulative adjective has been lemmatized in its own right, and this for two main reasons: (a) its high frequency, and (b) as a copulative adjective, it is homonymous with two other words — see (35).

- (35) **kuhle¹** ** *copulative cl. 15, cl. 17* < **okuhle** ► **it's good / beautiful / nice** ♦ Uku-lobola kuhle ngezizathu eziningi. • *Lobola is a good thing for various reasons.* ♦ Kuhle ukushada umuntu omthandayo. • *It's nice to marry someone you love.*
kuhle² ** *conjunction* ► **must; ought to** ♦ Zaqala ukweluleka izingane zazo zithi kuhle zilingise uGabha. • *They began to advise their children, saying they ought to imitate Gabha.*
kuhle³ *
 ▪ **kuhle kwa- / okwa-** *adverb* ► **(just) like; as** ♦ Bajamelana kuhle kwamaqhude amabili. • *They stared at each other just like two cocks do.*

The various forms (*kuhle*, *akukuhle*, *kwakuhle*, *kusekuhle*, and *kwakukuhle*) were sampled, and the frequencies redistributed as 1,944, 1,169 and 741 respectively. The use as a copulative adjective thus turns out to be the most frequent of the three. In comparison, a dictionary user who consults Dent and Nyembezi's dictionary, will only find '**kuhle** (adv) like' and '**kuhle** (conj) ought', in this order, while Doke and Vilakazi only treat the adverbial use. Both these existing dictionaries also fail to provide a crucial (encoding) feature, namely that as an adverb, *kuhle* is always followed by the PC17 *kwa-*, or the pronominalized indefinite PC15 *okwa-*. In our user-friendly dictionary, these are all provided for. A user who looks up the copulative use under *okuhle* (which is the 'normal' thing to do given the dictionaries' lemmatization policy), will be referred to **kuhle¹**: see the cross-reference before the first sense in (34), as well as the usage note at the bottom there.

11. Other words formed from adjective stems

Sections 3 to 10 introduced a new way to lemmatize adjectives in a user-friendly Zulu–English dictionary. Before we conclude, one last important point must be made. As has no doubt become clear from the discussion so far, words that belong together, no matter the size of the set, are best *treated together* — 'in one go', so to say. In this way one makes sure that one has truly considered everything that is common to each member, while highlighting what makes certain forms different from what is common — a variant of the well-known lexicographic tool *per genus proximum et differentia(e) specifica(e)*. Once one has completed this job, one must however also consider the wider picture, and treat all related forms. In the case of adjectives, a large number of words can be derived from the adjective stems, words that end up in other word classes. The Addendum shows all the 'derivations' belonging to the top 5 000 lemmas.

A total of 82 lemmas may be said to be linked to and derived from the adjective stems, five of which are not covered in any of the existing dictionaries for Zulu (these are marked in bold in the Addendum). The overall frequency for these 82 forms is about 100 000 (97 430 to be exact), so two-thirds of the overall frequency of the adjectives themselves. It is interesting to see that one only finds derivations with the frequent adjective stems (those with a tick (✓) in the Z-E column of Table 1), except for *ngamafuphi* 'in brief' (286), a 'new' word which may be analyzed as follows: instrumental formative *nga-* + adjective concord *ama-* (referring to *amagama* 'words') + adjective stem *-fuphi* 'short', or thus 'with short words'. (Note that all 'derivations' with *-nye* are derived from the enumerative stem *-nye*, rather than from the adjective stem *-nye*.)

12. Pros and cons of the user-friendly lemmatization of adjectives in Zulu

Bringing the various strands together, and polarizing the extremes first, one may imagine at one end of the spectrum a *purely stem-based* lemmatization approach to the Bantu languages, whereby only the smallest meaningful morphemes are lemmatized and used as entry points for all members of the lemma as well as for all 'derived' items. Applied to the adjectives that would mean lemmatizing *core adjective stems* only, and under each of these twenty-odd stems, one would not only provide detailed guidance on the various qualificative and copulative uses (as discussed in Section 7), but also list all adjectives with extensions (such as diminutives and augmentatives), as well as all (main) derivations (such as all the items with other parts of speech listed in the Addendum). An approach like this would result in massive articles, each several pages long, the contents of which would need to be hierarchically and logically structured, but for the linguist and all language enthusiasts, this presentation would likely be the most rewarding one.

At the other end of the spectrum, one may imagine a *purely word-based* lemmatization approach, whereby each and every orthographic word is entered 'as is' into the dictionary. This effort, too, would be massive, and for all conjunctively written languages simply impracticable. Although extremely user-friendly for any beginner or even anyone with no knowledge whatsoever of the language concerned, such an approach would of course not only be endlessly repetitive, but would also miss out on important generalizations.

These two extremes are but two poles on a continuum, of course. In reality, a 'traditional' stem-based approach to lemmatization such as Doke's 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. Figure 2 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.

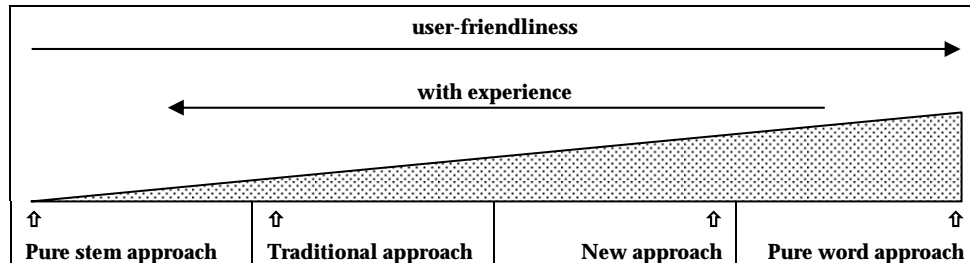


Figure 2: Stem versus word lemmatization for the Bantu languages

In the initial list of 20 000 items to be lemmatized (cf. Section 4), there were 332 adjectival forms. These were collapsed into 126 adjective articles — a move away from the 'pure word' pole, but still a *long* way from the 'pure stem' pole. Indeed, we settled for an approach that includes the adjective concord, as overall frequencies indicated that this form is also the most frequently used one. Note that of the 126 adjectives, about half (68) also have a star rating (cf. Tables 3 though 6: 17 x ***, 30 x **, 21 x *).

Given one is moving over a continuum, no matter which approach one settles for, there will always be pros and cons. The main '**cons**' of our new approach may be summarized as follows (with, between square brackets, a cross-reference to the relevant section where it was discussed above):

- Given the focus on top-frequent members only, none of the paradigms is ever complete. [4]
- For copulative adjectives, one needs to 'guess' the (abbreviated) relative concord. [6]
- For all adjectives with further prefixes, one needs to know or consult a (or 'the attached') grammar anyway. [7]
- Some of the (implicit) connections between words derived from the same adjective stem are lost. [11 and Addendum]
- One misses out on generalizations. [12]

In our view the '**pros**', which we list by way of conclusion below, far outweigh these few 'cons':

- Excellent reflection of the true distribution of the lexicon. [5]
- Precise translation equivalents are provided, rather than general ones. [5.1]
- The exact spelling for each form is given, without the need to apply the various morphophonological rules (useful for both receptive and active use). [5.2]
- The correct (and modern) class numbers are indicated, while class restrictions are implied. [5.3]

-
- Only the frequent core adjectives are treated, with (frequent) variant forms being cross-referenced to their more frequent forms. [5.4]
 - Diminutive and augmentative adjectives are listed directly there where they are also used. [5.1 and 5.2]
 - Typical combinations and collocations are entered where they are relevant. [5.3 (16) and 10 (34)]
 - Frequent derivations are listed there where they are relevant. [10 (34)]
 - Idiomatic use is pinpointed and covered there where it is relevant. [9]
 - Real examples illustrate each and every lemma, collocation, combination, derivation and idiomatic use. [5 through 10]
 - The most salient form of each lemma is illustrated, which in 90% of the cases is the lemma itself. [6]
 - The star rating (a logical by-product of the approach advocated here) gives a visual clue as to each adjective's relative importance. [2, 8 and 12]
 - The detailed analysis of corpus evidence also allows for a move towards the inclusion of supra-semantic features, such as the attention to semantic prosody. [5.1 (8)]
 - There is less dependency on a grammar for successful dictionary consultation. [5 through 10]
 - The ultimate user-friendliness is flexibility: for instance, when it comes to the need to differentiate between various homographs in a text, word-like adjectives may now be juxtaposed with words in other word classes through the inclusion of homonymous forms. [10 and -khulu in the Addendum]

It is our contention, then, that this new approach to the lemmatization of adjectives in Zulu will result in a more successful dictionary look-up experience.

Acknowledgements

Heartfelt thanks to my Zulu mentor, Prof. A. Wilkes. Thanks too, to the Publisher who was willing to embark on this innovative project, as well as to Ghent University for its continued support of my field trips to South Africa.

Endnotes

1. For an analysis of this dictionary, and its implications, cf. De Schryver and Wilkes (2008).
2. Comparing (2)(a) with (2)(b) one also notices that Doke assigned the part of speech (POS) 'noun' to *isine*/**izine*. This is incorrect, *isine* (at least when used neutrally) is an adverb. There is also no *izine* (corpus frequency = 0). Furthermore, in our user-friendly dictionary the form for 'fourthly' is lemmatized in its own right, under *okwesine*, which is the basic form from which Doke's *ngokwesine* is derived. Further note that the numbers between the (supposed) singular and plural forms in (2)(a), as well as in (1)(a), are tone markings, one per syllable.

3. Poulos and Msimang (1998: 142) list three more adjectives, namely: *-ngakanani* 'how big?', *-ngaka* 'so big, as big as this' and *-ngako* 'as big as that'. In Doke and Vilakazi (1953), these three stems are given both relative and adjective status. Given the adjective concords are different from the relative concords for classes 1 + 3, 4 and 6, corpus frequencies for adjectival forms in these specific classes can pinpoint whether or not these stems are indeed (frequently-used) adjectives. The results are: *omngakanani* (0), *emingakanani* (0), *amngakanani* (2); *omngaka* (1), *emingaka* (3), *amngaka* (2); *omngako* (0), *emingako* (0), *amngako* (0). Extrapolating from this, one can safely say that these three stems are *not* adjective stems. Conversely, Poulos and Msimang fail to mention the third most frequent adjective, *-ningi* (and its derivation *-ningana*), as well as *-ngaki* and *-fushane*. Although not a textbook, but a full-blown linguistic analysis, Poulos and Msimang erred in the same way as Taljaard and Bosch.
4. The 'proof' that this is a valid approach for Zulu lexicography will be given in a forthcoming article, but see De Schryver (2003) for a similar approach, applied to Ndebele.
5. For those not familiar with the numbering system for the Bantu noun classes: Classes 1, 3, 5, 7 and 9 contain singular nouns, with the corresponding plurals in classes 2, 4, 6, 8 and 10. Classes 11 and 14 contain mostly singular nouns, with some of them having plural features. In Zulu, classes 12 and 13 do not exist. Class 15 is the so-called infinitive class, containing (all the) verbs and (a few) lexicalized nouns. Classes 16 to 18 are the so-called locative classes which, for various Zulu parts of speech, can often be collapsed into just one class, class 17. This phenomenon is known as noun class reduction.
6. The various senses are always viewed from the Zulu point of view, which explains a series like 'small; young; little' under a single sense number, as one is dealing with a single concept in Zulu here.
7. These are, for the 1st person singular: *engimdala* (frequency = 4), *engimfushane* (1), *engimkhulu* (1), *engimncinyane* (1), *engimubi* (1) and *engimuhle* (1); for the 1st person plural: *esibabi* (2), *esibadala* (8), *esibahle* (1), *esibakhulu* (2), *esibancane* (1) and *esibaningi* (4); and for the 2nd person plural: *enibabi* (1), *enibadala* (3), *enibade* (1), *enibakhulu* (1) and *enibancane* (2).
8. The only other augmentative adjective that is frequent enough to be included is *omkhulukazi* (frequency = 53), with the same meaning as *enkulukazi* (56). *Eziningana* (101), mentioned in Section 5.1, is the only frequent diminutive adjective.
9. Cf. Endnote 5 for more on the Zulu classes.
10. Observe that this very paragraph is another example of this. Also, the hapaxes and other low frequencies, in Table 5 and elsewhere, are always 'suspect' — all class 15 forms in Table 5, for instance, come from textbooks only. The Bible is another source of many low-frequency words.
11. The frequencies of the adjectives *omfushane*, *omfishane* and *ezimfishane* is actually lower than 2 x 50, but these adjectives are nonetheless lemmatized for both classes in each case. Idem for *okudala* in Table 4.
12. If POS-tagged corpora were available for Zulu — as is for instance the case for Northern Sotho (De Schryver and De Pauw 2007) — the distribution for the different parts of speech would immediately be known. Sampling techniques would still be needed, however, and this (a) to pinpoint the relative distribution of the various senses for polysemous items, and (b) to make sure one has a truly mixed set of KWIC lines, randomly sampled from the various sources, for monosemous items. In general, some homonyms also have the same POSs, and for these sampling is also often a must.

References

- 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.** 2003. Drawing up the Macrostructure of a Nguni Dictionary, with Special Reference to isiNdebele. *South African Journal of African Languages* 23(1): 11-25.
- De Schryver, G.-M.** 2008. Why Does Africa Need Sinclair? *International Journal of Lexicography* 21(3): 267-291.
- De Schryver, G.-M.** 2008a. The Lexicographic Treatment of Quantitative Pronouns in Zulu. *Lexikos* 18: 92-105.
- De Schryver, G.-M. and G. De Pauw.** 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226-246.
- 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 E. Taljard.** 2007. Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho. *Lexikos* 17: 37-55.
- 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: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Doke, C.M., D.M. Malcolm and J.M.A. Sikakana.** 1958. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- Doke, C.M. and B.W. Vilakazi.** 1953² [1948]. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Fox, G.** 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London: Collins ELT.
- 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.
- Hanks, P.** 2002. Mapping Meaning onto Use. Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 156-198. EURALEX.
- Joffe, D. et al.** 2008. *TshwaneLex Suite* [online]. <http://tshwanedje.com/tshwanelex/>.
- Mbatha, M.O.** 2006. *Isichazamazwi sesiZulu*. Pietermaritzburg: New Dawn Publishers.
- Nyembezi, C.L.S.** 1992. *Isichazamazwi sanamuhla nangomuso*. Pietermaritzburg: Reach Out Publishers.
- Nyembezi, C.L.S. and O.E.H. Nxumalo.** 1966. *Inqolobane yesizwe*. Pietermaritzburg: Shuter & Shooter.
- Poulos, G. and C.T. Msimang.** 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.
- Scott, M.** 2008. *WordSmith Tools* [online]. <http://www.lexically.net/wordsmith/>.
- Sinclair, J.M.** 1966. Beginning the Study of Lexis. Bazell, C.E., J.C. Catford, M.A.K. Halliday and R.H. Robins (Eds.). 1966. *In Memory of J.R. Firth*: 410-430. London: Longman.
- Sinclair, J.M.** 1987. Collocation: A Progress Report. Steele, R. and T. Threadgold (Eds.). 1987. *Language Topics: Essays in Honour of Michael Halliday*: 319-331. Amsterdam: John Benjamins.
- Sinclair, J.M.** 1998. The Lexical Item. Weigand, E. (Ed.). 1998. *Contrastive Lexical Semantics*: 1-24. Current Issues in Linguistic Theory 171. Amsterdam: John Benjamins.
- Taljaard, P.C. and S.E. Bosch.** 1993² [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.

Addendum: All words, with a lemmatized corpus frequency ≥ 50 , 'derived' from adjective stems

- bi 'bad'** > adverb: *kabi* 'badly; very (much)' (5432); noun: *ububi* 'evil' (578); locative adverb: *ebubini* 'from the evil' (83)
- bili 'two'** > adverbs: *isibili* 'second' (3016), *kabili* 'two times' (1024), *ngambili* 'both' (63), *okwesibili* 'secondly; for the second time' (841); conjunction: *nambili* 'and two' (460); noun: *uLwesibili* 'Tuesday' (457); inclusive numeral pronouns (cf. De Schryver 2008a): *bobabili* 'both (of them)' (1445), *kokubili* 'both' (166), *nobabili* 'both of you' (164), *omabili* 'both (of them)' (214), *sobabili* 'both of us' (351), *womabili* 'both (of them)' (375), *yomibili* 'both (of them)' (168), *zombili* 'both (of them)' (943), *zozimbili* 'both (of them)' (74)
- dala 'old'** > adverb: *kudala* 'long ago' (993); noun: *ubudala* 'old; age' (184)
- de 'long'** > adverbs: *kade* 'long ago' (4338), *kakade* 'long ago' (193), *kude* 'far' (2331), *ngesikade* 'at long last' (193), *phakade* 'forever' (382); noun: *ubude* 'length; height; depth' (442)
- fuphi 'short'** > adverb: *ngamafuphi* 'in brief' (286)
- fushane 'short'** > adverb: *kafushane* 'shortly' (173)
- hlanu 'five'** > adverbs: *isihlanu* 'five; fifth' (909), *kahlanu* 'five times' (61), conjunction: *nanhlanu* 'and five' (178), noun: *uLwesihlanu* 'Friday' (1099)
- hle 'good'** > adverbs: *kahle* 'well; carefully' (20382), *kahlehle* 'very well; very much; precisely' (100); nouns: *isihle* 'kindness' (79), *ubuhle* 'goodness' (1231); locative adverb: *ebuhleni* 'near/in/... beauty' (62)
- khulu 'big'** > adverb: *kakhulu* 'very much' (19249); conjunction: *namakhulu* 'and hundreds' (117); nouns: *ikhulu/amakhulu* 'hundred/~s' (256/898) [in the plural a (more frequent) homonym of the class 6 adjective *amakhulu* 'big'], *indlunkulu* 'main hut; royal house' (133), *isikhulu/izikhulu* 'important person/~s' (1228/2050), *onkulunkulu* 'gods' (136), *ubabamkhulu/obabamkhulu* '(my/our) grandfather/~s' (77/80), *ubukhulu* 'greatness; size' (374), *umdlunkulu* 'chief's wife/wives' (67), *umkhulu/omkhulu* 'grandfather/~s' (330/83) [in the plural a (lesser frequent) homonym of the class 1or3 adjective *omkhulu* 'big'], *undlunkulu* 'member of the royal family' (65), *undunankulu* 'premier' (606), *uNkulunkulu* 'God' (4473), *uthishomkhulu* 'principal' (51), *uyisemkhulu/oyisemkhulu* '(her/his, their) grandfather/~s' (95/50); locative adverbs: *ekomkhulu* 'in/at/to/from/... the head office' (50), *endlunkulu* 'in/at/to/from/... the main hut; in/at/to/from/... the royal house' (237), *ezikhulwini* 'to/from/among/... important persons' (122), *komkhulu* 'in/at/to/from/... the chief/king's place' (1000)
- ncane 'small / few'** > adverb: *kancane* 'a little; slowly' (5572)
- ne 'four'** > adverbs: *isine* 'fourth' (493), *kane* 'four times' (73), *okwesine* 'fourthly; for the fourth time' (55); conjunction: *nane* 'and four' (182); noun: *uLwesine* 'Thursday' (352)
- ngaki 'how many?'** > adverb: *kangaki* 'how often?' (62)
- ningi 'much/many'** > adverb: *kaningi* 'many times' (656); nouns: *iningi* 'the majority' (2549), *ubuningi* 'abundance; plural' (734)
- sha 'new'** > adverb: *kabusha* 'anew' (805); noun: *intsha* 'youth' (1328); locative adverb: *entsheni* 'to/among/... the youth' (126)
- thathu 'three'** > adverbs: *isithathu* 'third' (1380), *kathathu* 'three times' (459), *okwesithathu* 'thirdly; for the third time' (251); conjunction: *nantathu* 'and three' (126); noun: *uLwesithathu* 'Wednesday' (608); inclusive numeral pronouns (cf. De Schryver 2008a): *bobathathu* 'all three (of them)' (209), *sobathathu* 'all three of us' (52), *zontathu* 'all three (of them)' (61)

The Lexicographic Treatment of Quantitative Pronouns in Zulu

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

Abstract: In Zulu, there are three kinds of quantitatives: inclusive, exclusive and numeral. For the lemmatization of these, even existing traditional dictionaries felt the need to move away from a pure 'stem' approach towards a 'word' approach. In a new Zulu-English dictionary project, this is not only confirmed, but is taken one step further with particular attention to the microstructure.

Keywords: LEXICOGRAPHY, DICTIONARY, BILINGUAL, CORPUS, LEMMATIZATION, FREQUENCY, ZULU (ISIZULU), ENGLISH, QUANTITATIVE PRONOUN, INCLUSIVE QUANTITATIVE PRONOUN, EXCLUSIVE QUANTITATIVE PRONOUN, (INCLUSIVE) NUMERAL QUANTITATIVE PRONOUN, USER-FRIENDLY

Samenvatting: **De lexicografische behandeling van kwantitatieve pronomina in Zoeloe.** In Zoeloe zijn er drie soorten kwantitatieven: inclusieve, exclusieve en numerieke. Voor de lemmatisatie daarvan voelden zelfs bestaande traditionele woordenboeken de nood om stam-lemmatisatie te verlaten voor woord-lemmatisatie. In een nieuw Zoeloe-Engels woordenboekproject wordt dit niet alleen bevestigd, doch wordt ook een stap verder gegaan, met bijzondere aandacht voor de microstructuur.

Sleutelwoorden: LEXICOGRAFIE, WOORDENBOEK, TWEETALIG, CORPUS, LEMMATISATIE, FREQUENTIE, ZOELOE, ENGELS, KWANTITATIEF PRONOMEN, INCLUSIEF KWANTITATIEF PRONOMEN, EXCLUSIEF KWANTITATIEF PRONOMEN, (INCLUSIEF) NUMERIEK KWANTITATIEF PRONOMEN, GEBRUIKSVRIENDELIJK

1. Distribution of Zulu parts of speech

This short contribution is one in a series of research articles that deal with the lemmatization of the different Zulu word classes, or parts of speech (POSs), in a user-friendly Zulu-English dictionary for junior users. In De Schryver and Wilkes (2008) the *possessive pronouns* were discussed, in De Schryver (2008a) an in-depth analysis of the *adjectives* followed, while the *quantitative pronouns* are the topic of the present endeavour.

The project proceeds within a Sinclairian, corpus-driven lexicographic framework (De Schryver 2008). For the Zulu to English side of the dictionary, a corpus of 8.5 million Zulu words was first lemmatized, and the top 5 000 lem-

mas — together with their lemmatized corpus frequencies — constitute the backbone of the dictionary's Zulu macrostructure. As it turns out, each Zulu lemma with a lemmatized corpus frequency of at least 50 needs to be considered for inclusion in the dictionary.

The three largest categories of Zulu parts of speech are — unsurprisingly — the different types of nouns, verbs and adverbs. Together, and in a dictionary that contains the top 5 000 lemmas only, these three POS categories cater for about 80% of the Zulu lexicon. We refer to these three as 'Group 1'. The other POS categories can be divided into two further groups based on the number of members in each POS category. 'Group 2' consists of those POS categories with around 100 or (slightly) more members — these are the relatives, adjectives, conjunctions, possessive pronouns, and ideophones. 'Group 3' consists of all the rest, thus POS categories with (much) fewer than 100 members each — these include the interjections, enumeratives, demonstrative pronouns, quantitative pronouns, relativized possessive pronouns, locative demonstrative copulatives, absolute pronouns, etc.

A detailed study of the top three POS categories, thus Group 1, will be undertaken in future contributions. Given two POS categories from Group 2 (viz. adjectives, with 126 members, and possessive pronouns, with 99 members) have already been looked into, it is now appropriate to briefly engage with one of the smaller POS categories from Group 3 — *in casu* the quantitative pronouns, with just 33 members. In dictionary terms, the three categories that will have been discussed so far amount to just 5.16% of the planned total of 5 000 dictionary articles, as illustrated in Figure 1.

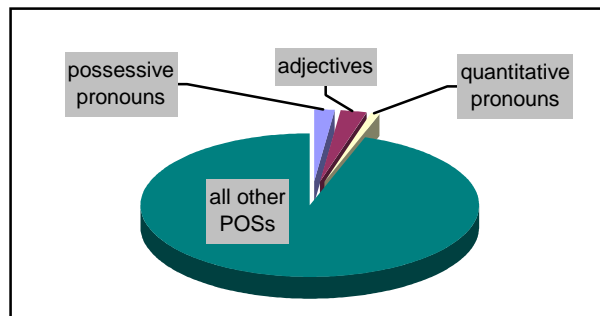


Figure 1: Zulu POS categories studied from a lexicographic point of view

The possessive pronouns amounted to 1.98% of the total, the adjectives to 2.52%, and the quantitative pronouns are good for 0.66%. There is no reason to believe that there is a correlation between the size of a particular POS category and the lexicographic difficulty of that category. Each POS category deserves a discussion in its own right, and once one will have covered all categories, cross-POS discussions will surely be required for it all to make even more sense. Nor is there a reason to believe that there is a strong correlation between the size of

a particular POS category and the summed frequency of its members. The final say on these aspects will only be possible near the end of the project, however. This said, the quantitative pronouns seem not to pose too many lexicographic problems. Yet neither are they trivial.

2. Zulu quantitative pronouns: A brief linguistic perspective

There are three types of quantitative pronouns in Zulu, all used to express *quantity*, viz. the inclusive quantitative pronouns, the exclusive quantitative pronouns, and the numeral quantitative pronouns. The inclusive quantitative stem is *-nke*, which means 'the whole' when referring to singular nouns and 'all' when referring to plural nouns. The exclusive quantitative stem is *-dwa*, for which the basic meaning is 'alone; only'. For the numeral quantitatives, any of the following adjective stems may be used: *-bili* 'two', *-thathu* 'three', *-ne* 'four', and *-hlanu* 'five'. The quantitative pronouns are formed as shown in Table 1.

Table 1: The formation of the inclusive, exclusive and numeral quantitative pronouns in Zulu

(with *Cl.* = noun class number and 1st and 2nd persons; *SC* = subject concord; *PR* = pronominal root; *QStem* = quantitative stem; *AP* = adjective prefix; *AStem* = adjective stem (illustrated for *-bili* only); *N* = nasal, i.e. n or m)

Cl. ↓	Inclusive quantitative pronouns				Exclusive quantitative pronouns				Numeral quantitative pronouns				
	SC	PR	QStem	↓	SC	PR	QStem	↓	SC	PR	AP	AStem	↓
1	u-	-o-	-nke	> wonke	u-	-o-	-dwa	> yedwa	—	—	—	—	—
2	ba-	-o-	-nke	> bonke	ba-	-o-	-dwa	> bodwa	ba-	-o-	ba-	-bili	> bobabili
3	u-	-o-	-nke	> wonke	u-	-o-	-dwa	> wodwa	—	—	—	—	—
4	i-	-o-	-nke	> yonke	i-	-o-	-dwa	> yodwa	i-	-o-	mi-	-bili	> yomibili
5	li-	-o-	-nke	> lonke	li-	-o-	-dwa	> lodwa	—	—	—	—	—
6	a-	-o-	-nke	> onke	a-	-o-	-dwa	> odwa	a-	-o-	ma-	-bili	> omabili
7	si-	-o-	-nke	> sonke	si-	-o-	-dwa	> sodwa	—	—	—	—	—
8	zi-	-o-	-nke	> zonke	zi-	-o-	-dwa	> zodwa	zi-	-o-	ziN-	-bili	> zozimbili
9	i-	-o-	-nke	> yonke	i-	-o-	-dwa	> yodwa	—	—	—	—	—
10	zi-	-o-	-nke	> zonke	zi-	-o-	-dwa	> zodwa	zi-	-o-	ziN-	-bili	> zozimbili
11	lu-	-o-	-nke	> lonke	lu-	-o-	-dwa	> lodwa	—	—	—	—	—
14	bu-	-o-	-nke	> bonke	bu-	-o-	-dwa	> bodwa	bu-	-o-	bu-	-bili	> bobubili
15	ku-	-o-	-nke	> konke	ku-	-o-	-dwa	> kodwa	ku-	-o-	ku-	-bili	> kokubili
17	ku-	-o-	-nke	> konke	ku-	-o-	-dwa	> kodwa	ku-	-o-	ku-	-bili	> kokubili
1sg	ngi-	-o-	-nke	> wonke	ngi-	-o-	-dwa	> ngedwa	—	—	—	—	—
1pl	si-	-o-	-nke	> sonke	si-	-o-	-dwa	> sodwa	si-	-o-	ba-	-bili	> sobabili
2sg	u-	-o-	-nke	> wonke	u-	-o-	-dwa	> wedwa	—	—	—	—	—
2pl	ni-	-o-	-nke	> nonke	ni-	-o-	-dwa	> nodwa	ni-	-o-	ba-	-bili	> nobabili

In Table 1, one sees that subject concords (SCs) consisting of a vowel only, change to their semivowel (*u* > *w*; *i* > *y*), while *a* is dropped. The vowel of the other SCs is elided. The 1st person singular of the inclusive quantitative takes the form of class 1, while class 1 as well as the 1st and 2nd persons sin-

gular of the exclusive quantitative is irregular. These four forms, which do not follow the pattern, have been marked in bold. The formation of only one of the numeral quantitatives is illustrated in Table 1, namely with the adjective stem *-bili* 'two'; the meaning of which becomes 'both'. The formation and meaning of the other three numeral quantitatives is similar. For classes 8 and 10 morpho-phonological rules apply: the nasal *N* is *m* before **b** (applies to *-billi*), *n* elsewhere; the combination **n** + **th** becomes **nt** (applies to *-thathu*).

The information presented so far is what one typically finds in textbooks and linguistic analyses of Zulu (cf. e.g. Taljaard and Bosch (1993: 83-85), or Poulos and Msimang (1998: 124-129)). Such sources will also list a few more features, some of which will be encountered below. Conversely, corpus evidence reveals other features which none of the existing sources mention. Before these can be discussed, we need to turn from linguistics to lexicography.

3. Moving from the 'stem pole' to the 'word pole' in lemmatizing quantitative pronouns

As is well known, the lemmatization policy adopted in all existing Zulu dictionaries is to group the lexicon around word stems. So, in a dictionary such as Doke and Vilakazi's (1953) *Zulu-English Dictionary*, a user is able to look up the six stems *-nke*, *-dwa*, *-bili*, *-thathu*, *-ne* and *-hlanu*. Under both the inclusive and exclusive quantitative stems, this user is even given all the full forms listed in Table 1. Surprisingly, the full forms themselves have *also* been lemmatized in addition. When it comes to the adjective stems *-bili*, *-thathu*, *-ne* and *-hlanu*, while not all full quantitative forms are listed within the articles of the stems, a cross-reference to the first (few) of the series is given. Here too, the full forms themselves have been lemmatized *in addition*. Clearly, then, this is a hybrid approach — one in which Doke and Vilakazi is simultaneously acting as a *stem* and *word* dictionary. If the stem and word approaches are viewed as two poles on a continuum (De Schryver 2008a: 86-87), then one could say that Doke and Vilakazi physically move about on this continuum in their dictionary. No doubt, this hybrid approach was followed for reasons of retrievability, or thus user-friendliness.

Interestingly, in Dent and Nyembezi's (1995) *Scholar's Zulu Dictionary*, a dictionary in which an attempt is made to make it easier to find words, the inclusive stem has not been lemmatized, while the exclusive stem, just like the adjective stems, has been. No list of all the forms may be found under the exclusive stem, however, while no further guidance at all with regard to the quantitative pronouns is given under the adjective stems. Yet again, the full forms (or at least, some of them, cf. the Addenda) have been lemmatized. It seems that, in an attempt to lower the threshold, the compilers of this dictionary moved even further away from the 'stem pole'.

Indeed, in a truly user-friendly dictionary there is little point in listing the six stems used in forming quantitative pronouns. This is even true for the

adjective stems when used as adjectives (De Schryver 2008a). In a user-friendly dictionary, one thus moves radically *away* from the 'stem pole', towards the 'word pole'. Given one is dealing with a continuum, the next obvious question is: 'Where to make the cut?' In other words, which formatives and/or prefixes does one keep for lemmatization? Here overall corpus frequencies quickly reveal that the forms as shown in Table 1 are also the ones that are best lemmatized. This will become clear in the discussion below.

4. Inclusive quantitative pronouns

Of the six quantitative pronouns, the inclusive quantitative pronoun is by far the most frequent, as may be seen from the data in Table 2.

Table 2: Distribution of the quantitative pronouns
(with *Freq.* = the summed frequency of all (lemmatized) forms)

Quantitative pronoun	Freq.	%
inclusive (- <i>nke</i>)	47 799	72.33
exclusive (- <i>dwa</i>)	13 700	20.73
2 (- <i>bili</i>)	3 931	5.95
3 (- <i>thathu</i>)	451	0.68
4 (- <i>ne</i>)	174	0.26
5 (- <i>hlanu</i>)	34	0.05
	66 089	100.00

The actual breakdown of the inclusive quantitative pronoun has been tabulated in Addendum 1, where the left side of the table summarizes the corpus statistics, and links these to the user-friendly Zulu–English dictionary under construction; while the right side shows the data as seen in Dent and Nyembezi's dictionary.

Clearly, all forms are frequent enough to be included in any user-friendly dictionary, which was consequently also done in both dictionaries. However, given Dent and Nyembezi do not indicate for which classes certain translation equivalents apply, there is considerable room for confusion. For *sonke*, for example, their equivalents are 'all of us; all of it'. A more user-friendly approach is (1).

- (1) **sonke** *** *inclusive quantitative pronoun 1 1p pl* ► **all of us** ♦ La manzi asetshenzi-swa yithina sonke lapha ekhaya. • *This water is used by all of us here at home.* **2 cl. 7** ► **the whole; each; every** ♦ Wabonakala efunda izincwadi zesikole sonke isikhathi. • *He was seen reading school books the whole time.*

In (1), sense numbers are used to present each different class in its own right, while corpus examples illustrate each class. The same applies, *mutatus mutandis*, to all other classes.

Poulos and Msimang (1998: 126) point out that 'it is ... not uncommon to hear people say' *wonke* rather than *onke* (and likewise for the other class 6 quan-

tatives). The use of a corpus enables one (a) to see whether or not this is also reflected in the orthography, (b) if it is, to see how (un)common it really is, and (c) to use the results during corpus-driven dictionary compilation.

The orthographic form *wonke* occurs a staggering 6 414 times in 8.5 million words, so it is obviously not feasible to read through all concordance lines. What one can easily do is to sample, and this is what is done in Figure 2 — where the software (*WordSmith Tools*, Scott 2008) has been instructed to randomly select every one-hundredth instance only.

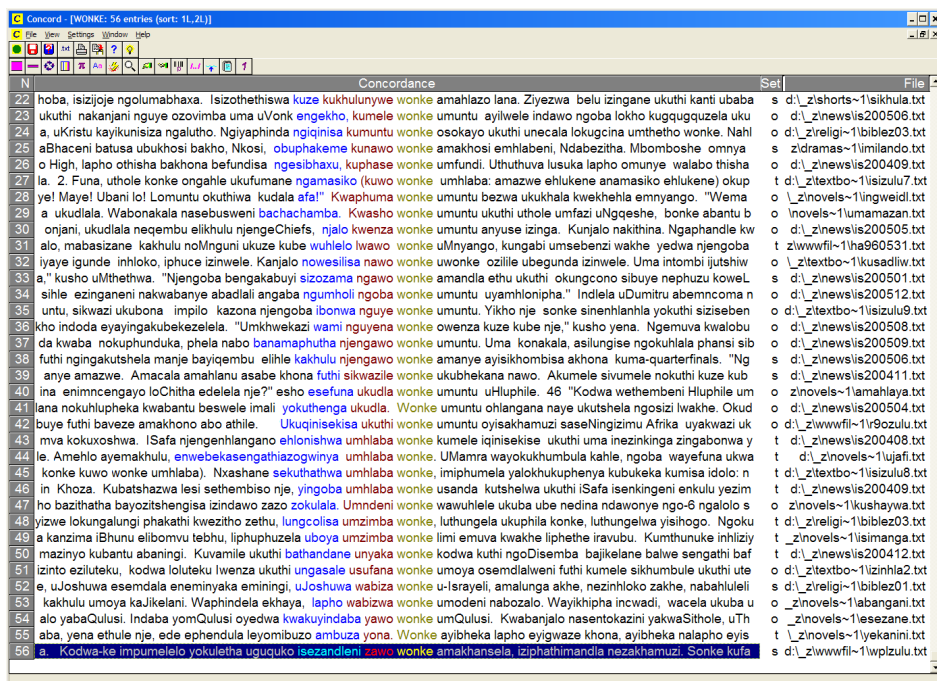


Figure 2: Sampling *wonke*, which occurs in several classes (in the screenshot, under 'Set', 'o' stands for class one, 't' for class three, and 's' for class six)

Rather surprisingly, and extrapolating from the sample, not only does *wonke* indeed occur in class 6, its frequency is as high as 1 489. The distribution across the different subcorpora (cf. the last column in Figure 2) is also even, with instances in short stories, dramas, newspapers, religious texts, etc. Furthermore, although grammars claim that the basic form for class 6 is *onke*, given the frequency of *onke* is 1 659, while that of *wonke* in this class is 1 489, it is clear that both forms are simply used interchangeably. This new information may be embedded into the respective dictionary articles. Compare (2) and (3).

- (2) **onke** ** *inclusive quantitative pronoun cl. 6* Compare **wonke** ► **all** ♦ Onke amehlo aphenduka abheka le moto. • *All eyes turned around and looked at this car.*
- (3) **wonke** *** *inclusive quantitative pronoun 1 cl. 1, cl. 3* ► **the whole; each; every** ♦ Lokhu kudalela umndeni wonke inkinga. Kuzokwenziwa njani manje? • *This caused a problem for the whole family. What is going to happen now?* ♦ Wonke umuntu owayelapho wamangala kabi. • *Each person who was there was very surprised.* 2 cl. 6 ► **all** ♦ Amandla wonke asemahlombe kaMnuz Bamba Ndwandwe. • *All the authority rests on the shoulders of Mr. Bamba Ndwandwe.* 3 1p sg ► **the whole of me** ♦ Sengiyibonile mina wonke. • *The whole of me has now seen it.* 4 2p sg ► **the whole of you** ♦ Abakithi bangilethele wena wonke ngogqoko. • *My friends brought the whole of you to me on a meat tray.*
- **wonke uwonke** cl. 1, cl. 3 ► **everyone; everybody** ♦ Wonke uwonke owayekhona lapho ngilanda le ndaba, wabamba ongezansi. • *Everybody who was there when I told the story was very surprised.*
 - ♦ **esewonke** cl. 6 ► **all together; the total** ♦ Esewonke amafulethi alapha angu-80. • *All together, there are 80 flats here.* ♦ Esewonke amaphepha okuvota angu-12 million. • *The total number of ballot papers is 12 million.*
- Note: For class 6, the pronoun 'wonke' also has the variant form 'onke', which is only slightly more frequent in this class.*

Note how a cross-reference and a usage note have been used in (2) and (3) respectively to bring all the information together. Further observe that frequent combinations (*wonke uwonke*) as well as derivations (*esewonke*) may all be treated under a single lemma such as *wonke*.

For the inclusive quantitatives, the latter is the exception rather than the norm, as for four of the nine inclusive quantitatives, the lemma sign is the *only* member of the paradigm. For the other five, (4) shows all the corpus forms that were brought together — during lemmatization — to obtain the lemma.

- (4) Lemma signs with members other than the lemma sign itself

konke <8769>	konke <8717>, konk <52>
sonke <3074>	sonke <3012>, sonkana <62>
wonke <6660>	wonke <6414>, esewonke <108>, wonkana <78>, uwonke <60>
yonke <7823>	yonke <7612>, yonkana <124>, isiyonke <87>
zonke <8815>	zonke <8577>, zonkana <74>, zonk <71>, sezizonke <50>, kuzozonke <43>

The forms without the final vowel are mostly found in poetry (in written Zulu; they *are* frequent in everyday speech), while those with the diminutive suffix *-ana* are used for extra emphasis. This leaves the instances in (5) to analyze.

- (5) Analysis of some of the forms from (4) (*with SC = subject concord*)

esewonke = SC6 in situative mood *e-* + auxiliary verb *-se* + SC6 in situative mood
e- + pronoun *wonke* = 'if they are now all together' (i.e. the sum/total, e.g. in exam papers) [cf. derivation under (3)]
 uwonke = SC3 *u-* + pronoun *wonke* = 'everyone' [cf. combination under (3)]

isiyonke = SC4or9 *i-* + auxiliary verb *-se* + SC4or9 in situative mood *i-* + pronoun *yonke* = 'it is now all/complete'
 sezizonke = auxiliary verb *-se* + SC8or10 in situative mood *zi-* + pronoun *zonke* = 'they are now all/complete'
 kuzozonke = locative prefix *ku-* + short form of absolute pronoun 'zona' *zo* + pronoun *zonke* = 'at/to/... all of them'

If one keeps a perspective on the various frequencies as seen in (4), however, then it is clear that the most productive way to lemmatize the inclusive quantitative pronouns is indeed under their basic forms. Low-frequency members of some of the paradigms, then, should only be illustrated when their meanings are lexicalized, as was the case in (3).

5. Exclusive quantitative pronouns

The corpus statistics and Dent and Nyembezi's treatment for the exclusive quantitative pronouns have been tabulated in Addendum 2. In addition to the data seen there, Dent and Nyembezi also lemmatized two diminutives: *yedwana* (corpus frequency = 109) and *yodwana* (43). These are indeed the two most frequent diminutive exclusive quantitative pronouns, but assigning them lemma-sign status does not seem warranted. Not tying certain translation equivalents to particular classes is again problematic as well. What they missed outright, and what even Doke and Vilakazi overlooked to cover explicitly, is an extra meaning which corpus data clearly reveals for all singular classes (1, 3, 5, 7, 9 and 11). Compare (6).

- (6) **lodwa** *** *exclusive quantitative pronoun cl. 5, cl. 11* ► **alone; on its own** ♦ Sibon'ikhwezi seliphezu kwentaba, Seliphume lodwa lathi qecelele [imidwa emi-2 yenkondlo] • *We saw the morning star above the mountain, It came out alone and stood out [2 poetry lines]* ♦ Lolu hlelo belungahambi lodwa njengoba belugcwele amaphutha. • *This programme could not proceed on its own because it was full of flaws.* ► **only** ♦ Wayegqoke ibhulukwe lodwa ngaphandle kweshethi. • *He was wearing trousers only, without a shirt.* ♦ Ungadla uphuthu lodwa uluqede. • *You can eat the thick porridge only and finish it.*
 ♦ **elilodwa** *cl. 5* ► **(only) one; (only) a single** ♦ Empeleni la maqembu asehlukaniswe yiphuzu elilodwa. • *In fact, these teams are now separated by only a single point.*
 ♦ **olulodwa** *cl. 11* ► **(only) one; (only) a single** ♦ Ezinye izichazamazwi zinolimi olulodwa vo, kanti ezinye zinezilimi ezimbili. • *Some dictionaries include only one language, while others include two languages.*

When a singular exclusive quantitative is preceded by a relative concord (RC) then the meaning becomes '(only) one; (only) a single', as may be seen from the article for *lodwa* in (6). *Lodwa* is actually an extreme example, as the frequencies of both *elilodwa* and *olulodwa* are higher than that of the lemma sign itself — see (7).

- (7) **lodwa** < 2270 > elilodwa <1206>, olulodwa <479>, lodwa <355>, nelilodwa <168>, lilodwa <62>

Clearly, then, it is *absolutely crucial* to use this information during the compilation of articles such as *lodwa*. During the project, the lexicographers are in the fortunate position to have the data shown in (7) at their disposal. Indeed, for each and every lemma and linked lemmatized corpus frequency, all the members of each paradigm (together with their individual frequencies) are available right there where they need it in *TshwaneLex* (Joffe et al. 2008), the dictionary writing system used.

For class 6, corpus data indicates that the 'variant form' *wodwa* (175) is actually slightly more frequent than what grammarians consider to be the basic form, *odwa* (163). The dictionary articles for *odwa* and *wodwa* may therefore be treated in a similar way as seen in (2) and (3).

Finally, if one extracts the various structures from each and every exclusive paradigm such as (7), one obtains all the possibilities listed in (8).

- (8) SC in indicative or situative mood + excl. pronoun
 RC + excl. pronoun
 excl. pronoun + diminutive *-ana*
 negative morpheme in indicative mood *a-* + negative SC1 in indicative mood *ka-* + excl. pronoun
 locative formative *ku-* + RC + excl. pronoun
 associative formative *na-* + RC + excl. pronoun
 instrumental formative *nga-* + excl. pronoun (+ diminutive *-ana*)

If one now considers the frequencies of each of the structures listed in (8) compared to the frequencies of the basic exclusive quantitative pronouns, then the statistics indicate that the latter is more frequent overall. This, then, is also why lemmatization was undertaken around the basic forms. In addition, for the two low-frequent exclusive quantitative pronouns, viz. *odwa* (163) and *nodwa* (73), the basic forms are also the only ones in the paradigm.

6. Numeral quantitative pronouns

The corpus and dictionary facts for the four numeral quantitative pronouns have been summarized in Addenda 3 to 6. Note the dramatic decrease in overall frequency going from 'both' to 'all five', viz. 3 931 → 451 → 174 → 34. Clearly, and right away, none of the forms listed in Addendum 6 ('all five') qualifies to be lemmatized within the top 5 000 Zulu lemmas, given the minimum lemmatized frequency is 50. Looking at the breakdown in Addendum 5 ('all four'), one concludes that none of these forms qualifies either.

This leaves us with only 'both' and 'all three'. For these two numeral quantitative pronouns, one immediately notices that Dent and Nyembezi overlooked to lemmatize the *most frequent* form in each case! These are *bobabili* 'both

(of them)' (cl. 2) and *bobathathu* 'all three (of them)' (cl. 2). This once again confirms why one needs a corpus rather than intuition in order to decide on what to include in and what to omit from a dictionary. As another example, *womabili* 'both (of them)' (cl. 6), has not been lemmatized, while the infrequent *yombili* has been. Two more points must be considered. Firstly, the frequency of *womabili* is 375, higher than that of *omabili*, which has a frequency of 214. Compare in this regard the full 'variant' status of the other class 6 quantitatives (*wonke/onke* and *wodwa/odwa*) discussed above. Secondly, *yombili*, with a frequency of just 18, is 'suspect'. In the corpus, it appears once in a textbook, once in the Bible, three times in novels, and 13 times in newspaper articles. The textbook example is exactly that: a textbook example. It was taken from Doke and Vilakazi's dictionary, *imfe yombili* 'both pieces of sweet corn', which thus quantifies a class 9 noun, a singular — this while all numeral quantitative pronouns are only *by definition* supposed to quantify items in the plural classes. Newspaper text is always suspect, and when in a huge text like the Bible only a hapax appears, one again has reason to doubt the status of that particular form. This leaves just three occurrences in novels, too few to make any linguistic claims, and far too few anyway to describe in a dictionary. (For completeness, the forms *yomthathu*, *yomne* and *yomhlanu* (luckily) do not occur in the corpus.)

Three of the four class 14 numeral quantitative pronouns do not occur at all, for 'all three', 'all four', and 'all five', while 'both' in class 14 has the lowest frequency of all forms in Addendum 3. Not treating any of these is thus the proper procedure. Corpus statistics further reveal that the so-called variant forms for classes 8 and 10 are actually more frequent than their basic forms: *zozimbili* (74) vs. *zombili* (943); *zozintathu* (14) vs. *zontathu* (61); *zozine* (30) vs. *zone* (30); *zozinhlanu* (2) vs. *zonhlanu* (7). This has direct implications for dictionary making, as the cross-reference must go from the lesser frequent to the most frequent form. Compare (9) and (10).

- (9) **zombili** ** *inclusive numeral pronoun cl. 8, cl. 10* ► **both (of them)** ♦ Akuzange kulimale muntu kuzo zombili lezi zehlakalo. • *Nobody was injured on both these occasions.* ♦ Undebembili: Leli gama lisho ukuthi izindebe zombili ziyasetshenziswa ekuphinyisweni komsindo. • *Bilabial: This word means that both lips are used in the pronunciation of a sound.*

- (10) **zozimbili** *inclusive numeral pronoun cl. 8, cl. 10* = **zombili**

Perhaps a note on the examples is necessary at this stage. Just as was the case for all other dictionary articles, whenever two different classes need to be exemplified in the main section of a dictionary article, an example for each has been selected. This was done, so that the dictionary is also a *didactic tool*, which conveys information both explicitly and implicitly. Further observe that the Zulu word for 'bilabial' (obviously) also includes the adjective stem *-bili* 'two' (*undebembili* (Cl. 1a/2a) < *izindebe* (Cl. 11/10) 'lips' + *mbili* (Cl. 10) 'two'). The frequency of *undebembili* in the general language is however too low to be lemmatized.

Further note that the POS in (9) and (10) is 'inclusive numeral pronoun', as the numeral quantitatives may indeed be seen as being derived from the inclusive quantitatives (for which the inclusive stem has been dropped) followed by the adjectives.

Lastly, the reason why the numeral quantitatives (or thus the inclusive numeral pronouns) have been lemmatized under their basic forms is simply because each of these forms is also the only member of the paradigm. Nothing else is pre- or suffixed here.

7. Sinclairian lexicography

Within a Sinclairian, corpus-driven approach to dictionary making, sound lexicographic decisions accompany every step of the compilation; from the use of a corpus for the construction of the macrostructure, which includes decisions on *how to lemmatize* each and every part of speech, all the way to a *detailed analysis of meaning* and the *presentation* thereof in a dictionary. This has been exemplified for the Zulu quantitative pronouns in this contribution.

Acknowledgements

Heartfelt thanks to my Zulu mentor, Prof. A. Wilkes. Thanks too, to the Publisher who was willing to embark on this innovative project, as well as to Ghent University for its continued support of my field trips to South Africa.

References

- Dent, G.R. and C.L.S. Nyembezi. 1995³ [1988², 1969]. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.
- 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. 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: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Doke, C.M. and B.W. Vilakazi. 1953² [1948]. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.
- Joffe, D. et al. 2008. *TshwaneLex Suite* [online]. <http://tshwanedje.com/tshwanelex/>.
- Poulos, G. and C.T. Msimang. 1998. *A Linguistic Analysis of Zulu*. Cape Town: Via Afrika.
- Scott, M. 2008. *WordSmith Tools* [online]. <http://www.lexically.net/wordsmith/>.
- Taljaard, P.C. and S.E. Bosch. 1993² [1988]. *Handbook of isiZulu*. Pretoria: J.L. van Schaik.

[Abbreviations used in the **Addenda**: **Z-E** = user-friendly Zulu-English dictionary; **Freq.** = lemmatized corpus frequency (in 8.5 million words); **Band** = frequency band in Z-E, with *** the top 500 lemmas, ** the next 500, and * the third 500; **Cl.** = noun class number and 1st and 2nd persons; **CP** = noun class prefix; **Lemma sign**, **POS** and **Translation equivalent(s)** as in Dent and Nyembezi's (1995) dictionary]

Addendum 1: Inclusive quantitative pronouns

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
✓	6 660	***	1	um(u)-	wonke	(pron)	the whole of it.
✓	7 859	***	2	aba-	bonke	(pron)	all; all of them.
(✓)	(6 660)	(***)	3	um(u)-	wonke	(pron)	the whole of it.
✓	7 823	***	4	imi-	yonke	(pron)	all of them; the whole of it.
✓	2 642	***	5	i(li)-	lonke	(pron)	all of it.
✓	1 659	**	6	ama-	onke	(pron)	all; the whole of.
(✓)	(6 660)	(***)	wonke	-	-
✓	3 074	***	7	isi-	sonke	(pron)	all of us; all of it.
✓	8 815	***	8	izi-	zonke	(pron)	all of them.
(✓)	(7 823)	(***)	9	iN-	yonke	(pron)	all of them; the whole of it.
(✓)	(8 815)	(***)	10	iziN-	zonke	(pron)	all of them.
(✓)	(2 642)	(***)	11	u(lu)-	lonke	(pron)	all of it.
(✓)	(7 859)	(***)	14	ubu-	bonke	(pron)	all; all of them.
✓	8 769	***	15	uku-	konke	(pron)	all of it; everything.
(✓)	(8 769)	(***)	17	uku-	konke	(pron)	all of it; everything.
(✓)	(6 660)	(***)	1p sg		wonke	(pron)	the whole of it.
(✓)	(3 074)	(***)	1p pl		sonke	(pron)	all of us; all of it.
(✓)	(6 660)	(***)	2p sg		wonke	(pron)	the whole of it.
✓	498	*	2p pl		nonke	(pron)	all of you.
47 799							

Addendum 2: Exclusive quantitative pronouns

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
✓	3 373	***	1	um(u)-	yedwa	(pron)	he/she alone.
✓	1 115	**	2	aba-	bodwa	(pron)	they alone.
✓	1 239	**	3	um(u)-	wodwa	(pron)	it alone.
✓	1 993	**	4	imi-	yodwa	(pron)	they alone; he alone; it alone.
✓	2 270	***	5	i(li)-	lodwa	(pron)	it alone.
✓	163	-	6	ama-	odwa	(pron)	they only; only them.
(✓)	(1 239)	(**)	wodwa	-	-
✓	972	**	7	isi-	sodwa	(pron)	we alone.
✓	383	-	8	izi-	zodwa	(pron)	they alone.
(✓)	(1 993)	(**)	9	iN-	yodwa	(pron)	they alone; he alone; it alone.
(✓)	(383)	(-)	10	iziN-	zodwa	(pron)	they alone.
(✓)	(2 270)	(***)	11	u(lu)-	lodwa	(pron)	it alone.
(✓)	(1 115)	(**)	14	ubu-	bodwa	(pron)	they alone.
✓	998	**	15	uku-	kodwa	(pron)	alone; only.
(✓)	(998)	(**)	17	uku-	kodwa	(pron)	alone; only.
✓	548	*	1p sg		ngedwa	(pron)	me alone.
(✓)	(972)	(**)	1p pl		sodwa	(pron)	we alone.
✓	573	*	2p sg		wedwa	(pron)	you alone.
✓	73	-	2p pl		nodwa	(pron)	you alone.
13 700							

Addendum 3: Numeral quantitative pronoun '2'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
✓	1 445	**	2	aba-	bobabili	-	-
—	—	—	3	um(u)-	—	—	—
✓	168	-	4	imi-	yomibili	(pron)	both.
-	18	—	yombili	(pron)	both.
—	—	—	5	i(li)-	—	—	—
✓	214	-	6	ama-	omabili	(pron)	(pron)
✓	375	-	womabili	-	-
—	—	—	7	isi-	—	—	—
✓	74	-	8	izi-	zozimbili	(pron)	both.
✓	943	**	zombili	(pron)	both.
—	—	—	9	iN-	—	—	—
(✓)	(74)	(-)	10	iziN-	zozimbili	(pron)	both.
(✓)	(943)	(**)	zombili	(pron)	both.
—	—	—	11	u(lu)-	—	—	—
-	13	-	14	ubu-	bobubili	-	-
✓	166	-	15	uku-	kokubili	(pron)	both.
(✓)	(166)	(-)	17	uku-	kokubili	(pron)	both.
—	—	—	1p sg	—	—	—	—
✓	351	-	1p pl	—	sobabili	(pron)	both of us.
—	—	—	2p sg	—	—	—	—
✓	164	-	2p pl	—	nobabili	(pron)	both of you.
3 931							

Addendum 4: Numeral quantitative pronoun '3'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
✓	209	-	2	aba-	bobathathu	-	-
—	—	—	3	um(u)-	—	—	—
-	15	-	4	imi-	yomithathu	(pron)	three of them.
—	—	—	5	i(li)-	—	—	—
-	27	-	6	ama-	omathathu	(pron)	all three.
-	27	-	womathathu	-	-
—	—	—	7	isi-	—	—	—
-	14	-	8	izi-	zozintathu	(pron)	all three of them.
✓	61	-	zontathu	(pron)	all three of them.
—	—	—	9	iN-	—	—	—
(-)	(14)	(-)	10	iziN-	zozintathu	(pron)	all three of them.
(✓)	(61)	(-)	zontathu	(pron)	all three of them.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	bobuthathu	-	-
-	15	-	15	uku-	kokuthathu	(pron)	all three.
(-)	(15)	(-)	17	uku-	kokuthathu	(pron)	all three.
—	—	—	1p sg	—	—	—	—
✓	52	-	1p pl	—	sobathathu	(pron)	all three of us.
—	—	—	2p sg	—	—	—	—
-	31	-	2p pl	—	nobathathu	(pron)	all three of you.
451							

Addendum 5: Numeral quantitative pronoun '4'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
-	29	-	2	aba-	bobane	-	-
—	—	—	3	um(u)-	—	—	—
-	13	-	4	imi-	yomine	(pron)	the four of them.
—	—	—	5	i(li)-	—	—	—
-	35	-	6	ama-	omane	(pron)	all four.
-	27	-	womane	-	-
—	—	—	7	isi-	—	—	—
-	30	-	8	izi-	zozine	(pron)	all four of them.
-	30	-	zone	(pron)	all four.
—	—	—	9	iN-	—	—	—
(-)	(30)	(-)	10	iziN-	zozine	(pron)	all four of them.
(-)	(30)	(-)	zone	(pron)	all four.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	bobune	-	-
-	2	-	15	uku-	kokune	(pron)	all four.
(-)	(2)	(-)	17	uku-	kokune	(pron)	all four.
—	—	—	1p sg	—	—	—	—
-	6	-	1p pl	—	sobane	(pron)	all four of us.
—	—	—	2p sg	—	—	—	—
-	2	-	2p pl	—	nobane	(pron)	all four of you.
174							

Addendum 6: Numeral quantitative pronoun '5'

Z-E	Freq.	Band	Cl.	CP	Lemma sign	POS	Translation equivalent(s)
—	—	—	1	um(u)-	—	—	—
-	11	-	2	aba-	bobahlanu	-	-
—	—	—	3	um(u)-	—	—	—
-	1	-	4	imi-	yomihlanu	(pron)	the five of them.
—	—	—	5	i(li)-	—	—	—
-	4	-	6	ama-	omahlanu	(pron)	all five.
-	1	-	womahlanu	-	-
—	—	—	7	isi-	—	—	—
-	2	-	8	izi-	zozinhlanu	(pron)	all five of them.
-	7	-	zonhlanu	(pron)	all five of them.
—	—	—	9	iN-	—	—	—
(-)	(2)	(-)	10	iziN-	zozinhlanu	(pron)	all five of them.
(-)	(7)	(-)	zonhlanu	(pron)	all five of them.
—	—	—	11	u(lu)-	—	—	—
-	0	-	14	ubu-	bobuhlanu	-	-
-	5	-	15	uku-	kokuhlanu	(pron)	all five.
(-)	(5)	(-)	17	uku-	kokuhlanu	(pron)	all five.
—	—	—	1p sg	—	—	—	—
-	1	-	1p pl	—	sobahlanu	(pron)	all five of us.
—	—	—	2p sg	—	—	—	—
-	2	-	2p pl	—	nobahlanu	(pron)	all five of you.
34							

Bilingual Dictionaries, the Lexicographer and the Translator

Rachéle Gauton, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (rachel.gauton@up.ac.za)*

Abstract: This article focuses on the problems, and advantages and disadvantages of the bilingual dictionary from both the lexicographer's and the translator's point of view, with specific reference to bilingual Zulu dictionaries. It is shown that there are many and varying problems the lexicographer has to deal with and take cognisance of when compiling a translation dictionary. Of these, the main problem is the basic lack of equivalence or anisomorphism which exists between languages. This non-equivalence between languages is also the root cause of the difficulties with which the translator or user of the bilingual dictionary has to contend. The problems experienced by translators therefore overlap to a great extent with those which the lexicographer experiences in compiling a bilingual dictionary. It is concluded that the user of a bilingual dictionary should not only know what to expect to find in a translation dictionary, but must also treat such a dictionary with caution and discernment. It is also shown that there are clear criteria which the lexicographer can follow in compiling a bilingual dictionary, which would then enable the user (and in particular the translator as user) to disambiguate the recorded information successfully.

Keywords: BILINGUAL DICTIONARIES, LEXICOGRAPHER, TRANSLATOR, ISIZULU, PROBLEMS EXPERIENCED BY LEXICOGRAPHERS, PROBLEMS EXPERIENCED BY TRANSLATORS, NONEQUIVALENCE BETWEEN LANGUAGES, CHARACTER, SHORTCOMINGS AND ADVANTAGES OF BILINGUAL DICTIONARIES

Opsomming: Tweetalige woordeboeke, die leksikograaf en die vertaler. Die fokus van hierdie artikel is op die probleme, en voor- en nadele van die tweetalige woordeboek vanuit die oogpunt van sowel die leksikograaf as die vertaler, met spesifieke verwysing na tweetalige Zuluwoordeboeke. Daar word aangetoon dat daar baie en uiteenlopende probleme is waaraan die leksikograaf moet aandag gee en waarvan hy/sy moet kennis neem by die samestelling van 'n vertalende woordeboek. Die vernaamste van hierdie probleme is die basiese gebrek aan ekwivalensie of anisomorfisme wat tussen tale bestaan. Hierdie nie-ekwivalensie tussen tale is ook die grondoorzaak van die moeilikhede waarmee die vertaler of gebruiker van die tweetalige woordeboek moet worstel. Die probleme ondervind deur vertalers oorvleuel dus tot 'n groot mate met dié wat die leksikograaf ondervind by die samestelling van 'n tweetalige woordeboek. Daar word tot die slotsom gekom dat die gebruiker van 'n tweetalige woordeboek nie net moet weet wat om van 'n vertalende woordeboek te verwag nie, maar ook so 'n woordeboek omsigtig en oordeelkundig moet benader. Daar word ook getoon dat daar duidelike kriteria bestaan wat die leksikograaf kan volg by die samestelling van 'n tweetalige woordeboek wat die gebruiker (en veral die vertaler as gebruiker) dan in staat sal stel om die opgetekende inligting suksesvol te interpreteer.

Slutelwoorde: TWEETALIGE WOORDEBOEKE, LEKSIKOGRAAF, VERTALER, ISIZULU, PROBLEME ERVAAR DEUR DIE LEKSIKOGRAAF, PROBLEME ERVAAR DEUR DIE VERTALER, NIE-EKWIVALENSIE TUSSEN TALE, AARD, GEBREKE EN VOORDELE VAN TWEETALIGE WOORDEBOEKE

1. Introduction

This article focuses on the problems, advantages and disadvantages of the bilingual dictionary from both the lexicographer's and the translator's point of view, with specific reference to bilingual Zulu dictionaries.

Manning (1990: 159) indicates that the bilingual dictionary is the translator's basic tool, and that it is the bridge that makes interlingual transfer possible. Pinchuck (1977: 223) warns, however, that the bilingual dictionary is an instrument that has to be used with circumspection and discernment. Pinchuck (1977: 231) further cautions:

The bilingual dictionary has a particular importance for the translator, but it is also a very dangerous tool. In general when a translator needs to resort to a dictionary to find an equivalent he will do better to consult a good monolingual dictionary in the SL (source language — RG) and, if necessary, one in the TL (target language — RG) as well. The bilingual dictionary appears to be a short cut and to save time, but only a perfect bilingual dictionary can really do this, and no bilingual dictionary is perfect.

Swanepoel (1989: 202-203) agrees that it is a misconception to assume that the general bilingual dictionary is sufficiently sophisticated to be an ideal aid for professional translators. It is merely a useful, albeit a limited, aid. Swanepoel argues that the bilingual dictionary is limited for the following two reasons:

- (a) It does not contain sufficient information for the user.
- (b) It cannot be a substitute for the user's competence in the SL and TL. The process of translation involves the user's total communicative competence, which also includes a grasp of the text's sociocultural context.

Swanepoel concludes that the bilingual dictionary is nothing more than an aid to professional translators in cases where their acquired knowledge of the TL is lacking.

In this article, the reasons for this state of affairs will be elucidated by indicating

- (a) which problems are experienced by the lexicographer in the compilation of the bilingual dictionary, with specific reference to Zulu; and
- (b) which problems are experienced by the translator when attempting to find suitable translation equivalents by consulting the bilingual dictionary.

2. Problems experienced by lexicographers

According to Nida (1958: 279), the semantic problems involved in bilingual dictionaries are different from, and also more complicated than those encountered in the compilation of monolingual dictionaries. The reason for this is that

whereas monolingual dictionaries are prepared for users who participate in and understand the culture being described, bilingual dictionaries describe a culture which differs in various proportions from that of the users.

There are many and varying problems the bilingual lexicographer has to deal with and take cognisance of when compiling a translation dictionary. Of these, the main problem would seem to be the basic lack of equivalence or anisomorphism which exists between languages, and this is the issue that will mainly be focused on in this article.

3. The lack of equivalence between languages

Writers such as Zgusta (1971), Pinchuck (1977), Nelson (1978), Al-Kasimi (1983), Swanepoel (1989), Mtuze (1990), Neubert (1992), Baker and Kaplan (1994), Gouws (1996) and Adamska-Salaciak (2006) point out that there is a basic lack of equivalence or anisomorphism between languages. Zgusta (1971: 296) argues that a lack of equivalence can be manifested by any component of the lexical meaning, in any degree and dimension.

Baker and Kaplan (1994: 7) contend that equivalence is nebulous in nature, and cannot be represented by way of neat translation equivalents as is found in some conventional bilingual dictionaries. They point out that in interlingual communication, the same word, even with the same 'meaning', will have different equivalents in different contexts. Equivalence, therefore, is context-bound. Baker and Kaplan (1994: 8) further indicate that the problem of non-equivalence extends beyond the lexical level to the syntactic level. Although a certain syntactic structure may have an exact equivalent in another language, the equivalent structure may serve a different rhetorical purpose.

Zgusta (1971: 296) points out that very few equivalent words with no polysemy in either the SL or the TL have precisely the same meaning. He mentions that in the majority of cases such words are defined scientific terms. Zgusta (1971: 312) further indicates that usually the lexical meaning of the unit in the TL is only partly identical with that of its counterpart in the SL. The term 'partial equivalent' is therefore more appropriate than the term 'equivalent'.

Zgusta (1971: 319), as well as Al-Kasimi (1983: 60), indicates that there are basically two types of translations of entries found in a bilingual dictionary. These two types are:

- (1) Translational / insertable equivalents, e.g.
Zulu **umhlangano** English **meeting / assembly**
- (2) Explanatory / descriptive equivalents, e.g.
Zulu **ibandla** English **assembly of men (for discussion, as to hear a trial or an announcement of the chief)**

The difference between these two types of equivalents is that whereas the translational equivalent, i.e. **assembly / meeting**, can immediately be inserted

into a TL sentence, the explanatory equivalent, i.e. **assembly of men**, cannot always be directly inserted into a TL sentence. Al-Kasimi (1983: 60) points out that a further distinction can be made between an 'explanatory equivalent' and an 'explanation'. Where an explanation tends to be similar to a definition or description, an explanatory equivalent tends to approximate a translational unit that might be standardised by acceptance and use in the language, e.g. **assembly of men (for discussion, as to hear a trial or an announcement of the chief)** (Doke et al. 1990: 64). In this example, 'assembly of men' is an explanatory equivalent, whilst the phrase in parenthesis constitutes an explanation.

Zgusta (1971: 319) indicates that the main characteristic of a translational equivalent is that it must have the ability to be used in a fluent, good translation of whole sentences, and therefore to be inserted into contexts of the TL. The explanatory equivalent, on the other hand, gives more information about the lexical unit of the TL. Zgusta (1971: 321) further points out that an explanatory equivalent makes considerable demands on the user's knowledge of the TL. Zgusta (1971: 322) states that it is the lexicographer's task to indicate the most general translational equivalents that have a broader range of application which, with their glosses, correspond to the multiple meaning of the lexicographical unit of the SL. He concludes that the explanatory and translational equivalents are therefore not as opposed as may be thought at first glance, although the translational equivalent is always a possible, and sometimes the best possible, choice for insertion into a real sentence. As translational equivalents can be directly employed, Al-Kasimi (1983: 61) indicates that such equivalents should be favoured in a bilingual dictionary aimed at speakers of the SL as an aid to produce the TL.

Al-Kasimi (1983: 61) asserts that an explanatory equivalent works well if the TL is the user's mother tongue, as such an explanatory equivalent may suggest to, or elicit from, the user another equivalent which fits the particular context.

Neubert (1992: 6) argues that translational equivalents should constitute not primarily translations, but rather carefully selected prototypical lexical patterns. They would consequently function as a type of springboard, supplying translational starting points from which adequate equivalents can be found. Users are therefore expected to put in the extra work to find the most suitable equivalent, based on their native competence. Neubert (1992: 12) states:

A prototype does not claim to be *the one and only* translation, that would be fiction, it is intended to lead the user to *the* translation that is textually compatible, i.e. that fits grammatically, lexically, stylistically, pragmatically. A prototype, then, is a kind of cognitive schema, a lexical model. It supplies a potential translation, *not* the typical or even the ideal translation. It gives the user a clear notion of one sense of a word without inundating him in notional complexity.

Baker and Kaplan (1994: 3) contend that translators rarely manage to insert dictionary equivalents into the context of what they are translating. They point out that a better dictionary would be one which provides information about

context and usage. Nelson (1978: 213) is in agreement with this sentiment when he argues that it would be better if lexicographers were to think contextually, rather than using simple one-to-one translations.

Gouws (1996: 16) also underscores the importance of context to support and supplement any translation equivalents supplied. He summarises the functions of the bilingual dictionary and bilingual lexicographer as follows:

Although bilingual dictionaries are employed as polyfunctional sources of semantic information, their main function is not a transfer of meaning. Bilingual dictionaries are aids in interlingual translations and have to focus on a treatment that enables the user to render a good and sound translation. The main aim of the dictionary should not only be the establishment of a relation of semantic equivalence between source and target language. Instead, a lexicographer has to endeavour to reach communicative equivalence.

3.1 Lack of translation equivalents in the TL

Both Zgusta (1971: 323) and Al-Kasimi (1983: 61) point out that a major problem the bilingual lexicographer has to contend with, is that the required equivalents cannot always be found in the TL. These writers identify the following instances where a language might not necessarily possess the required translational equivalents:

3.1.1 Lexical units with other than designative function

A lexical unit in the SL might not have a corresponding lexical unit in the TL. For example, the Zulu auxiliary verb stem **-thi** has only a grammatical function as carrier of tense, aspect, etc. when used together with an ideophone as in:

- (3) **Ngisathi shelele, ngizobuya masinyane nje** 'I'm just slipping out (quickly), I'll be back shortly'.

In this context (i.e. used together with the ideophone **shelele** 'of going for a short while / slipping out'), the SL item **-thi** has no lexical equivalent in the TL.

3.1.2 Culture-bound words

These words denote objects or concepts peculiar to, for example, the SL culture, which would mean that such culture-bound items would have no translational equivalents in the TL. For example, the Zulu verb **-lobola** is explained by Doke et al. (1990: 460) as 'supplement a marriage by the handing over of some present of goods or of an agreed number of cattle (or money in lieu thereof) on the part of the bridegroom's people to the father or guardian of the bride, in order to ensure the right of the bridegroom to any issue of the marriage'. In order to overcome this problem of a lack of translational equivalent, lexicographers utilise the explanatory equivalent 'pass (over the) lobolo' in their illustrative

sentences, where 'lobolo' is borrowed from Zulu **ilobolo** which is the corresponding noun formed from the verb stem **-lobola**. Mtuze (1990) illustrates that cultural issues could create problems for lexicographers because they might not comprehend certain concepts foreign to their own culture.

3.1.3 Onomasiological gaps

Apart from those words lacking translational equivalents as has been discussed here, so-called 'onomasiological gaps' in, for example, the TL are also caused by (scientific and technological) terminology in the SL which do not exist in the TL. Al-Kasimi (1983: 61) points out that a vocabulary can be expanded or extended in a number of ways as can be illustrated for Zulu by means of the following examples:

- (4) Word borrowing, e.g.
English **computer** Zulu **ikhompiyutha**
- (5) Coinage, e.g.
English **nuclear fission** Zulu **ukucanda ubuphakathi bento**
(Back-translation: 'to cleave the inside/nucleus of something' (Doke et al. 1990: 314))
- (6) Giving new meaning to existing words, e.g.
English **department (of education, etc.)** Zulu **umnyango (wemfundo, njll.)**
Note that the basic meaning of **umnyango** is 'door'.
- (7) Extending the meaning of existing words, e.g.
English **electricity** Zulu **ugesi**
The meaning of **ugesi** has been extended from the original meaning 'coal gas (as used for lighting and heat)' now also to include the meaning 'electric light, lamp; electricity'.
- (8) Compounding new words from existing elements from the language or from it and some other one, e.g.
English **counterfeit coin** Zulu **imalimbumbulu**
(< **imali** 'money' + **mbumbulu** 'counterfeit' (Doke et al. 1990: 480))

Apart from the lack of a translational equivalent in the TL as has been discussed here, there are also other instances of non-equivalence to be found between the SL and TL.

3.2 Instances of non-equivalence between languages

Al-Kasimi (1983: 63-67) identifies the following further instances of non-equivalence between languages that can all be illustrated with examples from Zulu as indicated in the paragraphs below.

Languages differ in their related grammatical categories. For example, in the Bantu language family, the grammatical category 'noun' contains a division of nouns into up to 23 different noun classes which each has its own class prefix, indicating mainly singular and plural, but also other semantic notions such as abstract, diminutive, augmentative, locative, etc. This is very different from the word category 'noun' as found in, for example, English.

Languages differ in their parts of speech. For example, a language such as Zulu distinguishes the word category 'ideophone' which does not exist in a language such as English.

A lexical unit in one language may not have a corresponding lexical unit in another language. In Zulu the interrogative marker **na?** has no lexical equivalent in a language such as English. This interrogative marker corresponds to interrogative sentences indicated by only a question mark in English, e.g.

- (9) **Uyasebenza na?** 'Are you working?'

Two languages may have different grammatical patterns to determine certain aspects of experience. A Zulu sentence such as **Uhambile** can mean either 'He/she is gone' or 'He/she has gone/left' in English. Some supplementary information is therefore needed in order to arrive at an adequate English translation.

One of two corresponding words in different languages may have undesirable connotations. In Zulu **umfazi** signifies not only '(married) woman', but also has the added connotation of 'wife'. It may also sometimes be used as a term of insult for a loose woman (Doke et al. 1990: 201). Because of this negative connotation in certain contexts, it should therefore be avoided as a general term for 'woman'.

A lexical unit in one language may have two or more components whilst its equivalent in another language may have only one component. For example, Zulu **Sala kahle** and **Hamba kahle** correspond to English **(Good)bye**, whilst Zulu **Sawubona** corresponds to English **Hello, Good day, Good morning, Good afternoon** and **Good evening**.

Two related items in two different languages may not cover the same semantic range as in the case of the Zulu word **umuthi** which represents both 'tree' and 'medicine' in English.

Substantives and adjectives in the TL may not always be considered as equivalents of the substantives and adjectives of the SL. Whereas Zulu nouns and so-called 'relative' adjectives derived from those nouns differ morphologically (as in example (10)(a) below), this is not always the case in English where the same word may function as noun and adjective (as in example (10)(b) below):

- (10) (a) Zulu **amanzi** (n) 'water', *but* **-manzi** (a) 'wet', e.g. **indwangu emanzi** 'a wet cloth'
 (b) English **stone** (n), *and* **stone** (a), e.g. **stone wall**

Idioms and figures of speech, i.e. exocentric expressions, create special difficulties for the bilingual lexicographer as a certain amount of adaptation is necessary for the translation of these expressions. A metaphor in the SL cannot simply be translated with a corresponding metaphor in the TL. For example, the English metaphor **Adam's apple** cannot be translated by way of a metaphor in Zulu, as such a metaphor does not exist. (The Zulu equivalent for 'Adam's apple' is the term **igilo**.) Mtuze (1990: 32) points out that idioms and other figurative expressions have become so fixed by usage in a particular language that it is very difficult or even impossible to render them in another language.

Swanepoel (1989: 216-217) indicates that languages also differ from each other regarding the presuppositions or connotations underlying lexical items. For example, the nearest translational equivalents for the Zulu word **umthakathi** is English 'witch, wizard, warlock'. There is, however, a significant difference between what is understood by the concept of practising witchcraft and the person who indulges in this type of practice within the Zulu culture, and the connotations associated with English 'witch, wizard, warlock'. Similarly, the translation of Zulu **isangoma** ('diviner') with English 'witchdoctor' ignores the fact that the practice of the diviner has nothing in common with witchcraft as either practised by the *umthakathi* within Zulu culture, or understood within Western culture. The connotations or presuppositions underlying the various words within the different cultures are therefore completely divergent.

Swanepoel (1989: 219) indicates that, because of problems such as have been discussed here regarding lack of translational equivalents, as well as the non-equivalence between languages, translators cannot reasonably expect the bilingual dictionary to provide them with translation equivalents that can be inserted directly into a TL text. Rather, the bilingual dictionary will provide translators with translational equivalents in a limited number of cases, and with explanatory equivalents and explanations in the majority of cases. It is furthermore essential that the bilingual dictionary should provide adequate examples illustrating the various contexts in which a specific word can be used.

Mtuze (1990: 32) points out that equivalence is 'a rather difficult ideal' for the lexicographer. Neubert (1992: 1) refers to the lexicographer's dilemma which 'consists simply in his brave attempt to do the impossible'. Mtuze (1990: 32) states:

At best, we try to maintain adequacy [...] by which we mean that the target language version is not 100% equivalent to the source language version but that it is, on the whole, a fair reflection of the original.

Adamska-Sałaciak (2006: 20) stresses the following inevitable paradox which renders bilingual dictionaries a contradiction in terms:

[...] bilingual dictionaries are impossible in theory, but indispensable and irreplaceable in practice.

From the preceding discussion, it is clear that there are many and varying problems the bilingual lexicographer has to deal with and take cognisance of

when compiling a translation dictionary. Of these, the main problem would seem to be the basic anisomorphism or lack of equivalence which exists between languages.

This non-equivalence between languages is also the root cause of the difficulties with which the translator or user of the bilingual dictionary has to contend.

4. Problems experienced by translators

The problems experienced by translators overlap to a great extent with those problems the lexicographer experiences in compiling a bilingual dictionary.

4.1 The nature of bilingual dictionaries — what the user should know

According to Pinchuck (1977: 223), the translator should bear in mind

- (a) that a dictionary, and therefore also a bilingual dictionary, is always out of date;
- (b) that many of the recorded expressions are no longer in common use;
- (c) that expressions referred to as colloquial or non-standard may have risen into more formal use; and
- (d) that, most commonly, new expressions have come into use but are not yet recorded.

Pinchuck points out that the dictionary therefore has limitations, but if used intelligently, it can be of great value, and indeed indispensable.

Neubert (1992: 1 *et seq.*) argues for a realistic attitude of the user towards the bilingual dictionary. Such a realistic attitude, Neubert believes, will be determined by the following factors:

- (a) Entries should be appreciated as constituting pieces of text which provide the user with *directions* on how to use target words as substitutes for source words.
- (b) The user should be aware of the 'default settings' of the dictionary, i.e. who the dictionary is aimed at / directed towards. Bilingual dictionaries cannot present the same information both ways, i.e. to L1 as well as to L2 speakers. The bilingual dictionary further constantly has to choose between either defining the meaning of an L1 item, or translating it by way of L2 material.
- (c) The translation(s) selected by the bilingual lexicographer as L2 equivalents should never be mistaken for *the* translation(s) of an L1 item. Such translational equivalents should rather be regarded as prototypical lexical patterns, directing the user to look in a certain direction for more

appropriate L2 equivalents. Neubert (1992: 6) states that this extra work on the part of the user is decisive. Neubert (1992: 7) rightly says:

The term translation dictionary does not entail that it offers the needed translation. It points the way to the translation, no more. Who thinks otherwise [...] takes fiction for fact.

As was also mentioned at the outset of this article, Swanepoel (1989: 203) states that the translation dictionary is merely a useful aid to professional translators in cases where their acquired knowledge of the second language is lacking. Swanepoel (1989: 203-204) further indicates that the translation dictionary does not immediately supply the most suitable equivalent for a certain context, but rather a collection of expressions in the TL from which the user must make a choice, or which may be used as a guideline in searching for the most suitable translation equivalent. Swanepoel points out that users of a translation dictionary must

- (a) know exactly what they are looking for in a translation dictionary, i.e. users must know how to interpret and evaluate the information given, and how to relate it to their specific usage tasks; and
- (b) know how to execute further searches in a specific translation dictionary and in other sources such as monolingual dictionaries in the TL and SL.

4.2 Shortcomings of the bilingual dictionary

Baker and Kaplan (1994: 2-3) indicate that dictionaries available to the translator tend to be found at opposite ends of the spectrum. At the one end, there are bilingual dictionaries which offer equivalent words and phrases, rather than explain the meanings of headwords. At the other end, there are the monolingual dictionaries which, irrespective of whether they are intended for mother tongue speakers or second language learners, tend to rely on synonyms. A new range of so-called 'semi-bilingual' dictionaries contain a combination of equivalents and synonyms. Baker and Kaplan are critical of the usefulness of dictionaries to the translator, as translators in real life work with genuine communicative events at the level of discourse, rather than with neat abstractions. Translators therefore rarely manage to insert dictionary equivalents into the context of what they are translating. As was mentioned earlier, Baker and Kaplan believe that a better dictionary is one that provides information about context and usage. This sentiment is echoed by Gouws (1996: 16-17):

Where the specific contexts in which translation equivalents can be used to substitute the lemma are not given as part of the lexicographical treatment, it is hardly possible that the creation of semantic equivalence can lead to the establishment of communicative equivalence. [...] Lack of additional information impedes the possibility to reach communicative equivalence; the form of equivalence that should be the lexicographer's first priority.

Pinchuck (1977: 225) indicates that ideally a dictionary should be an instrument of semantic discrimination which enables the user to choose between words for a given application. A dictionary should also offer information about the position of a word within a series and its value within a lexical structure. He argues, however, that this function is performed crudely in existing dictionaries, mainly because of the arbitrary arrangement of items from the linguistic and conceptual points of view. He contends that the dictionary presents the language as an inventory — as a list of words unconnected with one another, instead of offering it as a structured and patterned system, which is the way words are used in practice.

Pinchuck (1977: 232) argues that in a bilingual dictionary, neither the definition nor the single word equivalent meets the needs of translation. Should the dictionary provide long definitions, they are frustrating to the user who wants a word. Should the dictionary provide a single translation equivalent, it would also not be satisfactory, as it is bound to be a selection from a series of possible alternatives. Such a selection would then have been made according to the arbitrary judgement of the compiler. The ideal would therefore be the mapping out of the whole area of signification of the words of the language, by means of a series of equivalents. Such equivalents will not only serve to make the meanings and usage clear, but will provide or suggest the exact or most appropriate translation for the context in which the user has seen or heard a word or phrase.

Pinchuck (1977: 233) further points out that the bilingual dictionary depends for its comprehension on the maximum possible co-operation of the user, and relies on users' understanding of their mother tongue. Pinchuck therefore concludes that, in the light of these factors, the bilingual dictionary should only be used as a last resort.

Despite the shortcomings of the bilingual dictionary as listed in this section, bilingual dictionaries can be of considerable use to the translator provided that

- (a) the translator knows what to expect and what not to expect from a translation dictionary; and
- (b) the dictionary consulted is a good one that meets with certain set requirements.

4.3 Advantages of the bilingual dictionary

To summarise, Pinchuck (1977: 234) lists the main features of a good bilingual dictionary as follows:

- (a) The dictionary should provide correct translation equivalents, despite the difficulties inherent in bilingual lexicography as discussed so far in this article.

- (b) The dictionary should furnish as wide a range of application as possible for each item — the range will never be wide enough.
- (c) The dictionary should detail full grammatical information regarding word class, inflectional and derivational forms, and syntactic restrictions and applications.
- (d) The dictionary should give the level of usage of the equivalents provided.

As mentioned before, Baker and Kaplan (1994) are critical of existing bilingual dictionaries. Baker and Kaplan (1994: 8) further indicate that because of the basic lack of equivalence or anisomorphism existing between languages, an SL word or expression may not have a straightforward equivalent in the TL. They argue that such a state of affairs reflects the realities of translation where space restrictions and the need to communicate the message succinctly, rather than by way of a long exposition, are important considerations in almost any kind of translational activity.

Baker and Kaplan argue in favour of a new type of bilingual dictionary, the so-called 'bridge bilinguals'. According to Baker and Kaplan (1994: 1), bridge bilinguals are translated versions of monolingual dictionaries in which the explanation is translated into the TL / the user's mother tongue. Baker and Kaplan (1994: 3) point out that the advantage of this type of dictionary is that it provides far more information about the behaviour and meaning of the defined item than would be found in a conventional bilingual dictionary. Baker and Kaplan (1994: 5) indicate that in this type of dictionary, the use of equivalents to replace translated explanations are avoided as a rule, except where an explanation would irritate the user as a ready equivalent would provide the same information in a compact and more accessible form. Baker and Kaplan (1994: 6) give the following criteria for deciding when to replace an explanation with an equivalent:

- (a) The equivalent must be a high frequency word in the TL.
- (b) The TL equivalent cannot be a polysemous word; it must have one meaning only.
- (c) The TL equivalent should replace the headword in a translated version of the example(s) which follow the explanation.
- (d) The TL equivalent should not be a loanword with a different meaning in the SL.
- (e) The TL equivalent should have the same use as the headword.

Baker and Kaplan (1994: 9) indicate that bridge bilinguals are not only helpful as tools for language learning, but that they are also useful to professional translators as they provide translations of examples. Bridge bilinguals can be regarded as genuine translation assignments that are either undertaken by professional translators, or at least monitored and assessed by them. Teams en-

gaged in writing such bridge bilinguals, work to the same kind of brief as most translators, which entails that they

- (a) are commissioned by a client;
- (b) have a specific group of prospective users in mind;
- (c) are aware of the function the target text is meant to fulfil; and
- (d) adhere to fairly strict deadlines.

5. Conclusion

In this article, the various problems confronting the bilingual lexicographer as well as the translator using bilingual dictionaries have been detailed and discussed. It is clear that underlying all of these problems is the basic lack of equivalence or anisomorphism between languages. It is also evident that users of bilingual dictionaries should not only know what they can expect to find in a translation dictionary, but must also treat such a dictionary with circumspection and discernment. It was also shown that there are clear criteria which the lexicographer can follow in compiling a bilingual dictionary, which would then enable the user (and specifically the translator as user) to disambiguate the recorded information with great success.

References

- Adamska-Salaciak, A.** 2006. *Meaning and the Bilingual Dictionary. The Case of English and Polish.* Frankfurt am Main: Peter Lang.
- Al-Kasimi, A.M.** 1983. *Linguistics and Bilingual Dictionaries.* Leyden: E.J. Brill.
- Baker, M. and R. Kaplan.** 1994. Translated! A New Breed of Bilingual Dictionaries. *Babel* 40(1): 1-11.
- Doke, C.M., D.M. Malcolm, J.M.A. Sikakana and B.W. Vilakazi.** 1990. *English-Zulu Zulu-English Dictionary.* Johannesburg: Witwatersrand University Press.
- Gouws, R.H.** 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. *Lexikos* 6: 14-31.
- Manning, A.** 1990. The Role of Dictionaries and Context in the Translation Process: The Canadian Connection. Thelen, M. and B. Lewandowska-Tomaszczyk. *Translation and Meaning. Part 1:* 59-166. Maastricht: Euroterm.
- Mtuze, P.** 1990. Problems of Equivalence and Adequacy in a Trilingual Context. *South African Journal of Linguistics* 8(1): 30-32.
- Nelson, R.J.** 1978. Translation and Translating Dictionaries. *The Incorporated Linguist* 17(2): 36-37.
- Neubert, A.** 1992. Fact and Fiction of the Bilingual Dictionary. *EURALEX '90 Proceedings. IV International Congress:* 29-42. Barcelona: Bibliograf.
- Nida, E.A.** 1958. Analysis of Meaning and Dictionary Making. *International Journal of American Linguistics* 24: 279-292.
- Pinchuck, I.** 1977. *Scientific and Technical Translation.* London: Deutsch.
- Swanepoel, P.H.** 1989. *Only Study Guide for LEKPER-Q.* Pretoria: University of South Africa.
- Zgusta, L.** 1971. *Manual of Lexicography.* The Hague/Paris: Mouton.

Standardization or Stigmatization? Challenges Confronting Lexicography and Terminography in Sesotho sa Leboa*

V.M. Mojela, *Sesotho sa Leboa National Lexicography Unit, University of Limpopo, Turfloop Campus, Polokwane, Republic of South Africa*
(mojelav@ul.ac.za)

Abstract: The article investigates the effects of the underutilization of the abundant vocabulary of Sesotho sa Leboa, which results from a one-sided standardization approach owing to the disregard and stigmatization of most dialects. Sesotho sa Leboa has several dialects differing greatly concerning terminology, pronunciation and vocabulary. The situation is complicated and aggravated by a standardization which has sidelined more than half of the dialects because of factors such as the influence of colonialism and lack of government co-ordination, the missionary activities, and the influence of early writers and publications. The strict and narrow standardization of Sesotho sa Leboa resulted in the exclusion of large parts of the dialectal vocabulary, the forcing of dialect speakers to accept a foreign standard language, the creation of 'prestige' and 'inferior' dialects, the separation of the standard language from its own dialects and the awakening of a nationalistic spirit among some dialect-speaking communities. Finally solutions for the challenges caused by these developments are afforded.

Keywords: COLONIALISM, CORPUS, DIALECT, INFERIOR DIALECTS, LEMMATIZATION, LEXICOGRAPHY, MISSIONARY ACTIVITIES, NATIONALISM, ORTHOGRAPHY, PRESTIGE DIALECTS, STANDARD LANGUAGE, STANDARD DICTIONARY, SIDELINED DIALECTS, STIGMATIZED DIALECTS, TERMINOGRAPHY

Opsomming: Standaardisasie of stigmatisasie? Uitdagings waarvoor leksikografie en terminografie in Sesotho sa Leboa te staan kom. Die artikel ondersoek die uitwerking van die onderbenutting van die ryk woordeskat van Sesotho sa Leboa wat volg uit 'n eensydige standaardisasiebenadering deur die miskenning en stigmatisasie van die meeste dialekte. Sesotho sa Leboa het verskeie dialekte wat sterk verskil ten opsigte van terminologie, uitspraak en woordeskat. Die situasie word gekompliseer en vererger deur 'n standaardisasie wat meer as die helfte van die dialekte geslyl het vanweë faktore soos die invloed van kolonialisme en die ontbreking van regeringsamewerking, die sendingaktiwiteite, en die invloed van vroeë skrywers en publikasies. Die streng en eng standaardisasie van Sesotho se Leboa het gelei

* This article was presented as a keynote address at the Twelfth International Conference of the African Association for Lexicography, organized by the Tshwane University of Technology in collaboration with the IsiNdebele National Lexicography Unit, at the Soshanguve Campus, Pretoria, Republic of South Africa, 27–29 June 2007.

tot die uitsluiting van groot dele van die dialektiese woordeskat, die dwang van dialektspreekers om 'n vreemde standaardtaal te aanvaar, die totstandkoming van "invloedryke" en "ondergeskikte" dialekte, die skeiding van die standaardtaal van sy eie dialekte, en die opwekking van 'n nasionalistiese gees by sommige van die dialektspreekende gemeenskappe. Laastens word oplossings aangebied vir die uitdagings veroorsaak deur hierdie ontwikkelinge.

Sleutelwoorde: KOLONIALISME, KORPUS, DIALEK, ONDERGESKIKTE DIALEKTE, LEMMATISERING, LEKSIKOGRAFIE, SENDINGAKTIWITEITE, NASIONALISME, ORTOGRAFIE, INVLOEDRYKE DIALEKTE, STANDAARDTAAL, STANDAARDWOORDEBOEK, GESYLYNDE DIALEKTE, GESTIGMATISEERDE DIALEKTE, TERMINOGRAFIE

1. Introduction

Sesotho sa Leboa or Northern Sotho is one of the 11 official languages of the Republic of South Africa. It consists of around 27 dialects (Mokgokong 1966: 8-9), with the following among the major ones: Sekone, Sepedi, Seroka, Selobedu, Sepulana, Seřlokwa, Sekopa, Sehananwa, Sekgaga and Sephalaborwa. Only a few of these dialects are represented in the official standard language, i.e. Sekone, Sepedi, Sekopa, Sekgaga (of Mphahlele), and the dialects around Turfloop and Chuenespoort. The majority of the dialects contributed very little to the development of standard Sesotho sa Leboa with regard to vocabulary and structure. The dialects which were sidelined by standardization include, among others, dialects such as Seroka, Selobedu, Sepulana, Seřlokwa, Sehananwa, Sekgaga (of Maake) and Sephalaborwa, which border with the Vatsonga-Machangane and the Venda communities in the Lowveld and the northern part of the Limpopo Province. The authorities who initially developed standard Sesotho sa Leboa did not have an interest in these dialects because they regarded them to be too 'inferior' and too 'primitive' for use in developing the standard language. The reason for this was that these dialects were regarded as 'corrupt' versions of the 'prestige' dialects and standard Sesotho sa Leboa. The type of attitude which was adopted by the standardizing authorities towards these sidelined dialects is comparable to the following description of dialect by Allen and Linn (1986: 220):

This represents perfectly the idea of dialect and most of the connotations that surround the idea: a dialect is uncouth — an ugly, imperfect, corrupt version of the language which I myself speak perfectly.

2. Factors which facilitated the stigmatization of these dialects

The stigmatization of the majority of the dialects of Sesotho sa Leboa can be ascribed to factors such as the influence of colonialism and lack of government co-ordination, the missionary activities and their role in the standardization of Sesotho sa Leboa, and the influence of early writers and publications.

2.1 The influence of colonialism and lack of government co-ordination

The governments which ruled over South Africa during the late 19th and the early 20th centuries, i.e. at first those of the British in the Cape Colony and Natal, and the two Boer Republics of the Transvaal and the Orange Free State, and subsequently that of the Union after 1910, paid little attention to the development of education and the development of the languages of the indigenous South African communities.

Unlike the colonial rule which intervened to guide the standardization of languages like Kiswahili in the former East African Federation (Kenya, Uganda and Tanzania), and Shona in the former Rhodesia, the politicians in South Africa left the responsibilities of developing the indigenous education and the indigenous languages to the European missionaries. When explaining the influential role political authorities have in the standardization of languages, Mojela (1999: 20) cites the following example from the formation and the standardization of Shona:

It was the Government of Southern Rhodesia which took the initiative to appoint Clement Doke in 1929 to investigate and report to the Legislative Assembly on the possibility of unifying the dialects of the territory into a literary form for official and educational purposes and the standardization of the orthography.

It was only after 1929 that the Transvaal Education Department (TED) started making attempts at standardizing the Sotho languages in the former Transvaal which eventually led to the formation of the Language Boards (Mojela 2005: 46). In South Africa, for instance, it was only after the passing of the Bantu Education Act in October 1953 (Act No. 47 of 1953) that the South African government took over control of formal education from the missionaries. Even though the Bantu Education Act gave government power to establish the Language Boards to develop the indigenous languages, the missionary legacies still dominated the system, because the elite group who came to dominate the membership of the Language Committees and the Language Boards were still the same products of missionary education. As such, the Language Committees and the Language Boards continued to promote the dialects and the languages favoured by missionary policies.

2.2.1 The missionary activities

The development of the Northern Sotho orthography and the origin of its standard form were very much influenced by the work of the German missionaries in Sekhukhuneland. Karl Endemann (1836–1919) of the Berlin Evangelical Missionary Society came to South Africa in 1860 and established several missionary stations in Bopedi and Maleuskop between 1862 and 1867. In 1865 Endemann founded the Botšhabelo Missionary Station on the banks of the Olifants River (near the present Loskop Dam) which became an important education

centre for the Basotho ba Leboa communities (Mojela 1997: 13). His colleagues and successors included missionaries like Alexander Merensky, Hermann Wangemann, Heinrich Grützner, and Albert Nachtigal. The first Sesotho sa Leboa dialects the German missionaries learned to speak and convert to writing were the Sekopa and Sepedi dialects. Other Northern Sotho dialects did not have this missionary advantage since their areas had little, or no missionary activity.

These missionaries started developing orthographies in Sepedi and Sekopa, and translating the Bible and many religious publications into these dialects. These publications were meant to be used for the education of and communication with the indigenous Bapedi and Bakopa communities in order to teach them the Word of God. The Botšhabelo mission station, for instance, became an important education centre which provided for the development of 'Sepedi'.

The Sepedi publications produced through translations of biblical material and early missionary research were to be used as references in the future standardization of the language. The German missionaries in Bopedi did not regard Sepedi as a dialect of any language because they were unaware of Seroka, Selobedu, Sepulana, Seřlokwa, Sehananwa and Sephalaborwa. Many parents from areas such as Mapulaneng, Bolobedu and Bořokwa sent their children to study at Botšhabelo where they were taught in Sepedi.

Even though few missionary stations were later established in other parts of the former Transvaal, they did not pay much attention to the development of dialects, but relied on the publications produced in Bopedi. Mission stations like Medingen (Mmidinyene) in Bolobedu and Metz (Mmetse) in Makhutšwe did little to develop Sesotho sa Leboa dialects in their areas.

2.2.2 The role of the missionary activities in the standardization of Sesotho sa Leboa

The missionaries promoted the dialect or dialects in the areas where they operated.

The first Sesotho sa Leboa orthography by the German missionaries was in the Sepedi dialect. This elevated Sepedi to a superior status. The sidelined Sesotho sa Leboa dialects did not have these missionary orthographies, proving what Allen and Linn (1986: 218) say about dialect:

The idea that a language for which there exists no written form, a language which has not yet been alphabetized, is for that reason intrinsically inferior, not a real language, but a mere dialect.

The establishment of schools, tertiary institutions and health facilities at the missionary stations created 'mini-urban centres' in the rural areas which came to be regarded as places of civilization. Therefore, the dialects spoken in these areas became representative of civilization. Botšhabelo was one of these mini-

urban centres for the Bapedi communities.

Moreover, orthographies, publications and written materials were produced at these mission stations, and these would influence the future course of standardization in favour of the dialects of the communities in the vicinities of the missionary stations.

Lack of missionary activities in the Lowveld and the northern part of the former Transvaal compelled parents there to send their children to places such as Botšhabelo to receive education, thereby helping to promote the Sepedi and Sekopa dialects at the expense of their own dialects. After graduating from Botšhabelo, most of these young 'elite' group returned to places like Bolovedu, Bořlokwa, Senwabarwana, Makhutšwe (or Makhutšu) and Mapulaneng, speaking the 'language of civilization', i.e. Sepedi. Using Sepedi was to them a status symbol, while their own dialects were associated with illiteracy and inferiority.

As a result of the missionary legacy, the real standardization of Sesotho sa Leboa, which started during the 20th century, tended to be dominated by the Sepedi and Sekopa dialects — a dominance which sidelined most of the Sesotho sa Leboa dialects and which later came to be challenged by the Sekone dialect.

2.3 The influence of early writers and publications

Most of the young graduates from education centres like Botšhabelo, Emma-
rentia Geldenhuis Secondary School, Kilnerton College and Bethesda, who were not Bapedi, started abandoning the Botšhabelo orthographies and shifting to Sekone and the dialects around the present Polokwane and Mokopane. The pressure exerted by Sepedi was gradually reduced by the rise of early scholars and writers like Dr M.J. Madiba and other new generation writers who gradually distanced themselves from the Botšhabelo orthographies. Dr Madiba, who was himself of Ndebele origin, published the famous Mahlontebe series of school readers which was prescribed in primary schools from substandard A to standard 6. Just like Dr A.S.V. Barnes's Afrikaans and English series, i.e. 'Die Môreson-reeks' and 'The New Graded English Series', Madiba's Mahlontebe series dominated Northern Sotho Education for almost a quarter of a century since the early sixties of the 20th century.

The Mahlontebe series was based on the Sekone dialect and, from the mid 20th century, these booklets practically came to represent standard Northern Sotho, especially after the newly established Northern Sotho Language Board started compiling an official orthography and standardizing the language basing it on Madiba's orthography. The introduction of the Northern Sotho Language Board, for instance, came at a time when there were several missionary orthographies in Sepedi, Sekopa and Sekone, with little, or no written forms in the majority of the dialects of Sesotho sa Leboa.

When standardizing this language the Language Board used existing missionary and secular orthographies, the latter being those resulting from the

work of scholars like Dr M.J. Madiba and Mr O.K. Matsepe. In a special research study made by translating a one-page paragraph into a few major dialects of Sesotho sa Leboa, the results revealed a major reduction in the role of Sepedi in the official standard language approved by the first Northern Sotho Language Board. The dialectal contribution to standard Sesotho sa Leboa today can be analyzed as follows (citing only a few major dialects):

Sekone	50%	Sepedi	30%
Selobedu	2%	Seḽlokwa	8%
Seroka	2%	Sepulana	1%

The remaining dialects constitute 7%.

3. The standardization resulted in a stigmatization

Rather than being the development of a standard language, the standardization of Sesotho sa Leboa was more a 'cleaning and purifying of the language', a 'keeping the language clean and pure by sifting out inferior and dirty elements from the dialects' and a 'purifying of the language by excluding impure and corrupt versions'. The consequences of this strict and narrow type of standardization for Sesotho sa Leboa are, among others, the following:

3.1 The exclusion of large parts of the dialectal vocabulary

Almost two-thirds of the vocabulary of the language is stigmatized, being excluded from the standardized version. The rich vocabularies of the sidelined dialects which were supposed to be included in dictionaries and used in literary works are omitted from the standard language because (it is assumed) this inferior vocabulary will spoil the purity of the language. For instance, an extensive part of the vocabulary of Seroka, Selobedu, Sepulana, Seḽlokwa, Sehanawa and Sephalaborwa is excluded from the standard language. To those who developed Sesotho sa Leboa the inclusion of these dialects into the standard language seems to have been a mistake. Thus, in order to keep the standard language as pure as possible, these 'inferior' dialects had to be excluded from the language. Before 1994, there was not a single scholar who showed interest in developing orthographies for these stigmatized dialects — not even the speakers of these dialects themselves.

3.2 The forcing of dialectal speakers to accept a foreign standard language

The majority of the dialect-speaking communities are forced to accept the standard language which is not only practically foreign, but also totally different from their mother dialects. This proved to be a provocation to the spirit of

nationalism and a danger to the stability of the Sesotho sa Leboa standard language. The communities from the sidelined dialects were made to believe that their dialects are inferior to those which are much closer to the standard language. Most of the speakers of these dialects were, and still are wary of speaking their 'languages' in public for fear of being stigmatized themselves. The standard Sesotho sa Leboa, which is taught in schools in, among others, Bolobedu, Bořlokwa, Bokgaga, Senwabarwana and Mapulaneng, differ completely from the indigenous dialects spoken in these areas. In fact, the Sesotho sa Leboa standard language is more of a second language than a mother tongue to these communities.

3.3 The creation of 'prestige' and 'inferior' dialects

The prestige dialects are the dialects which were consulted or used to develop the standard language. This led to the elevation of these dialects to a superior status while the status of the dialects which were sidelined by standardization remain inferior, being stigmatized as low-class dialects. The speakers of the prestige dialects see no need of knowing or speaking the inferior dialects because (they assume) they will gain nothing from these dialects. On the other hand, the communities whose dialects have been downgraded to an inferior position, strive to know, understand, read and write the standard language, including the dialects closer to the standard language in order to elevate their status. Of course, this is a natural phenomenon, i.e. that people whose language has status or prestige speak only their language, while people whose language has lower or inferior status or prestige speak all the languages. In South Africa, for instance, almost all the indigenous communities can speak English, while from the English communities only a tiny section is able to utter even a phrase in one of the nine indigenous languages, the reason being that the English do not see any need of knowing the 'inferior' indigenous languages, while on the other hand the indigenous communities want to associate themselves with the status of English. This is also the case with the prestige dialects versus the inferior, low-class, stigmatized dialects of Sesotho sa Leboa. In most cases, the speakers of the prestige dialects of Sesotho sa Leboa would need an interpreter to understand the inferior dialects, while the speakers of these dialects usually communicate effectively in all the dialects of Sesotho sa Leboa. The Balobedu people understand all that is said in Sepedi, Sekone, Sekopa and all the dialects of Sesotho sa Leboa, while the Bakone and Bapedi will need interpreters to understand what the Mapulana and Balobedu say. Most speakers of the prestige dialects of Sesotho sa Leboa will need to have the following Lobedu expression explained:

- *Mola khaawe o ya khata* (Lobedu)
- *Moja sa gagwe o ya iphihla* (standard Sesotho sa Leboa)

- Literal meaning: a person who eats his/her food hides him-/herself, a person who eats his/her food doesn't want to be seen
- Meaning: you shouldn't tell other people about your fortunes

3.4 The separation of standard Sesotho sa Leboa from its own dialects

Standard Sesotho sa Leboa is much closer to Setswana and the Sesotho languages than to dialects like Seroka, Selobedu, Sepulana, Seřlokwa and Sephalaborwa. The following comparison of Selobedu (a Sesotho sa Leboa dialect), Setswana (an official language) and standard Sesotho sa Leboa, explains this fact:

Sesotho sa Leboa	Setswana	Selobedu	English
<i>mopani</i>	<i>nato/mopani</i>	<i>mořanare</i>	mopani tree
<i>leribiři</i>	<i>lerubisi</i>	<i>mmankhořo</i>	owl
<i>mmankgagane</i>	<i>mmamanthane</i>	<i>molema</i>	bat
<i>hlapi</i>	<i>tlhapi</i>	<i>khobe</i>	fish
<i>betha/itiya</i>	<i>betsa</i>	<i>mořa/tiya</i>	wallop
<i>legotlo</i>	<i>legotlo</i>	<i>lehořo/peba/mantoro</i>	mouse
<i>legapu</i>	<i>legapu</i>	<i>lesalabu</i>	watermelon
<i>nona</i>	<i>nona</i>	<i>kholophana</i>	be fat/gain weight
<i>mogaditswane</i>	<i>mogaditswane</i>	<i>mphekwa</i>	lizard
<i>bogobe</i>	<i>bogobe</i>	<i>booswa</i>	porridge
<i>kota</i>	<i>kota</i>	<i>mothate</i>	wooden pole
<i>bolela</i>	<i>bua</i>	<i>apa/bolabola</i>	talk
<i>natefa</i>	<i>natefa</i>	<i>řefa</i>	sweet/tasty
<i>molete</i>	<i>mosima</i>	<i>moina</i>	hole

These examples prove that Setswana can to a greater extent be considered a dialect of Sesotho sa Leboa (and vice versa) than Selobedu. This is also true of the Sesotho language. As such, the standardization of Sesotho sa Leboa has not only created a huge difference between the standard language and its own dialects, but has also brought it closer to other independent national languages which were supposed to be further apart from its dialects.

3.5 The awakening of a nationalistic spirit among the Balobedu and Mapulana communities

The one-sided standardization of Sesotho sa Leboa did not only stigmatize the majority of its dialects, but also encouraged the rise of a nationalistic feeling among the speakers of these so-called inferior dialects. The appearance of the name 'Sepedi' instead of 'Sesotho sa Leboa' in the 1996 Constitution of the Republic of South Africa (section 6.1) gave an impetus to the Balobedu and Mapulana to start demanding official status for their 'languages' (dialects).

These communities developed a spirit implying that if Sepedi can gain official status, so can Selobedu and Sepulana. The reaction from the Balobedu emerged immediately after the publication of the Constitution in 1996, as shown in the following statement of Archbishop Prince Madlakadlaka, chairperson of the Khelobedu Project in the *Sowetan* of 11 January 2007:

He said that the Khelovedu project started in 1996 after children complained that teachers prohibited them from speaking or writing in their home language. 'We launched the Khelovedu project, (which includes) members of the Modjadji royal family and the parents of the pupils.'

3.5.1 The developments among the Mapulana communities

The Mapulana communities, just like the Balobedu, are developing an orthography for their 'language' Sepulana. They started questioning the validity of the incorporation of Sepulana into standard Sesotho sa Leboa, which is now referred to as Sepedi in the 1996 Constitution. Knowing that Sepedi, just like Sepulana, is one of the dialects of Sesotho sa Leboa, the Mapulana people found it unacceptable that, by using the name 'Sepedi' for 'Sesotho sa Leboa', their 'language' Sepulana is now considered a dialect of Sepedi. In 2002, the Sepulana Language Development Committee wrote a submission to the Joint Constitutional Review Committee requesting a constitutional amendment which gives official status to Sepulana. The Joint Constitutional Review Committee transferred its submission to PanSALB for investigation, which in turn gave this assignment to the Sesotho sa Leboa National Language Body. This language body found the main cause for this demand to be the sidelining and the 'stigmatization' of the Sepulana dialect and the replacement of the name of the standard language Sesotho sa Leboa by Sepedi in the new Constitution of the Republic of South Africa. Consequently, the Sesotho sa Leboa National Language Body recommended to PanSALB to seek the reinstatement of the name 'Sesotho sa Leboa' in order to reunite the Basotho ba Leboa communities.

3.5.2 The reaction of the Balobedu community after 1994

About the Balobedu regarding their language, Boshego (2002: 1) says:

Some of the Balobedu (Valobedu) suggested that their language, Khelovedu, should not only be given due consideration but be included as an official South African language. They also suggested that it should be used as a medium of instruction in their schools.

However, these suggestions were not taken into consideration. Instead Selobedu together with Sekone, Seroka, Sepulana, Seřlokwa, Sekopa, Sehananwa, Sekgaga and Sephalaborwa was, according to Section 6.1 (Founding Provisions) of the Constitution of the Republic of South Africa, 1996, to become dia-

lects of Sepedi (Boshago 2002: 1). When the then President of the Republic of South Africa, Dr Nelson Mandela, visited Her Majesty Queen Modjadji V at Kheřakoni (the Balobedu Palace) on 23 February 1999 she requested him to consider the recognition of Selobedu as 12th official language. The *Sowetan* of 24 February 1999 reported as follows in this regard:

The Queen told Mandela that her language, Selobedu, was not officially recognized and her people were forced to be taught Sepedi, which is not their language.

The Balobedu, under the influence of the ideas of the Kara Heritage Institute of Dr Mathole Motshekga and the Balobedu community leaders like Archbishop Prince Madlakadlaka, are demanding a withdrawal from standard Sesotho sa Leboa (or 'Sepedi' as it is called in the Constitution) and the recognition of Selobedu (or 'Khelovedu' as they themselves call it) as official language. Archbishop Prince Madlakadlaka and the Balobedu communities have already written several submissions and memoranda to the Constitutional Court, the various Government Departments and PanSALB, demanding recognition of official status for their 'language' Khelovedu. At the beginning of 2007, Archbishop Prince Madlakadlaka was quoted in an article entitled 'Another Official Language' in the *Sowetan* of 11 January 2007, referring to 'Sepedi' (the name used by the Department of Education for 'Sesotho sa Leboa') as too foreign to the Balobedu learners. To the Archbishop, Sepedi is the main cause for the high failure rate of Balobedu learners in schools. When writing examinations, Balobedu learners are compelled to translate from their mother tongue (Selobedu) into a foreign language (Sepedi).

The *Sowetan* of 11 January 2007 also quotes a Lobedu learner who has failed the standard 10 examination, as saying:

I failed because I could not write or speak Sepedi very well. I could not finish writing my scripts during the exams. This is because I was translating Khelovedu into Sepedi, which takes more time. Thanks to the Department (of Education) and PanSALB for making it difficult for us to make it in school.

4. The challenges for Sesotho sa Leboa lexicography

As point of departure, the Sesotho sa Leboa National Lexicography Unit started writing the *Comprehensive Monolingual Dictionary* and the *Bilingual Sesotho sa Leboa/English Dictionary* in 2001, basing the research on the Sesotho sa Leboa Corpus which was compiled at the University of Pretoria under the leadership of the lexicographer and metalexicographer Prof D.J. Prinsloo, head of the Department of African Languages. The first volumes of these two dictionaries have already been published. The first volume of the *Sesotho sa Leboa/English Bilingual Dictionary* appeared with Maskew Miller Longman in May 2006, while the first volume of the *Comprehensive Monolingual Sesotho sa Leboa*

Dictionary was published by Nutrend Publishers in 2007. The fact that these newly published dictionaries are based on the already established corpus means that the stigmatized lexical items from the sidelined Sesotho sa Leboa dialects did not form part of the lemmatization in these dictionaries. The Sesotho sa Leboa lexicographers are faced with the major challenge of bridging the gap between the standard language and the sidelined, stigmatized dialects in order to make the standard language acceptable to all the communities by:

- (a) doing a thorough research into all the dialects, especially the sidelined dialects, in order to have a record of the complete potential vocabulary of Sesotho sa Leboa,
- (b) including all the stigmatized lexical items from the sidelined, stigmatized dialects of Sesotho sa Leboa in the already established corpus,
- (c) lemmatizing and subsequently standardizing the lexical items from all the stigmatized dialects in order to make Sesotho sa Leboa acceptable to all the Basotho ba Leboa communities, and
- (d) lemmatizing and subsequently including these lemmata in the dictionaries as variants for most of the standard lexical items of Sesotho sa Leboa.

5. Conclusion

It is important to realize that the lemmatization of the vocabulary of the stigmatized dialects of Sesotho sa Leboa will not only bridge the gap between the prestige dialects and the stigmatized dialects, but will also guarantee the unity and the stability of Sesotho sa Leboa which was endangered during the 20th century by the language authorities. The lemmatization and the subsequent standardization of the stigmatized dialects will not only unite the Basotho ba Leboa communities under one standard language, but will also increase the size of the lexicon of Sesotho sa Leboa extensively.

References

- Allen, H.B. and M.D. Linn (Eds.). 1986. *Dialect and Language Variation*. London/Montreal: Academic Press/Harcourt Brace Jovanovich Publishers.
- Boshogo, P.L. 2002. *Is Khelovedu a Language or a Dialect?* PanSALB Occasional Papers 8. PanSALB.
- Mojela, V.M. 1997. *The Development of Bantu Linguistics*. NSOT300 Study Guide. Pietersburg: University of the North.
- Mojela, V.M. 1999. *Prestige Terminology and its Consequences in the Development of Northern Sotho Vocabulary*. Unpublished D.Litt. et Phil. Thesis. Pretoria: University of South Africa.
- Mojela, V.M. 2005. Standardization and the Development of Orthography in Sesotho sa Leboa — A Historical Overview. Webb, V. *The Standardization of African Languages in South Africa*. Pretoria: University of Pretoria.

Mokgokong, P.C. 1966. *A Dialect-geographical Survey of the Phonology of the Northern Sotho Area*. Unpublished M.A. Dissertation. Pretoria: University of South Africa.

Newspapers

Sowetan, 24 February 1999.

Sowetan, 11 January 2007.

Constitution, Reports and Acts

The Constitution of the Republic of South Africa, Act 108 of 1996.

Lexicography and Linguistic Creativity*

Rosamund Moon, *Department of English, University of Birmingham, Birmingham, United Kingdom (r.e.moon.1@bham.ac.uk)*

Abstract: Conventionally, dictionaries present information about institutionalized words, phrases, and senses of words; more creative formations and usages are generally ignored. Yet text and corpus data provide ample evidence of creativity in language, showing that it is part of ordinary linguistic behaviour and indeed often systematic.

This article looks at four specific types of lexical creativity in English: figurative meaning, word formation, idioms, and spelling. Focusing on selected examples, it discusses corpus evidence and then treatment in (principally) three recent monolingual dictionaries for learners of English. It argues that, even taking into account the pedagogical function and limited scope of these dictionaries, more could be said about creative aspects of lexis, and the systematicity of creative usage. This would be of benefit and interest to dictionary users, and empower them.

Keywords: AFFIXATION, CORPUS, CREATIVITY, DICTIONARY COVERAGE, ENGLISH, FIGURATIVE LANGUAGE, IDIOMS, LEXICOGRAPHY, NEOLOGISM, NORMATIVENESS, SPELLING, WORD FORMATION

Opsomming: Leksikografie en taalkreatiwiteit. Normaalweg verskaf woordeboeke inligting oor geïnstutionaliseerde woorde, frases en betekenis van woorde; meer kreatiewe vorming en gebruike word gewoonlik geïgnoreer. Tog bied teks- en korpusgegewens volop bewyse van kreatiwiteit in taal, wat toon dat dit deel van gewone taalkundige gedrag is en inderdaad dikwels sistematies.

Hierdie artikel beskou vier spesifieke soorte leksikale kreatiwiteit in Engels: figuurlike betekenis, woordvorming, idioome, en spelling. Deur op uitgesoekte voorbeelde te fokus, bespreek dit korpusbewyse en daarna behandeling in (hoofsaaklik) drie resente eentalige woordeboeke vir aanleerders van Engels. Dit voer aan dat, selfs al word die opvoedkundige funksie en beperkte omvang van hierdie woordeboeke in ag geneem, meer gesê sou kon word oor die kreatiewe aspekte van leksis, en die sistematiese van kreatiewe gebruik. Dit sal tot voordeel en belang van woordeboekgebruikers wees, en hulle bemagtig.

Sleutelwoorde: AFFIGERING, KORPUS, KREATIWITEIT, WOORDEBOEKDEKKING, ENGELS, FIGUURLIKE TAAL, IDIOME, LEKSIKOGRAFIE, NEOLOGISME, NORMATIWITEIT, SPELLING, WOORDVORMING

* An earlier version of this article was presented as a keynote lecture at the Twelfth International Conference of the African Association for Lexicography, organized by the Tshwane University of Technology in collaboration with the IsiNdebele National Lexicography Unit, at the Soshanguve Campus, Pretoria, Republic of South Africa, 27–29 June 2007.

1. Introduction

It could be argued that lexicography has little business with linguistic creativity. Most dictionaries are essentially normative, and even if they aim at description rather than prescription, their object of description is generally the standard form of a language or dialect. As a result, creative phenomena such as metaphor, neologism, wordplay, and other exploitations are only considered to be part of the lexicon, and therefore worth including in a dictionary, when evidence suggests that a usage recurs and has become institutionalized.

Creative uses of language have, of course, long been discussed within linguistics: for example, in the contexts of humour (Nash 1985, Chiaro 1992), word coinage (Bauer 1983, Katamba 1994), and spoken interaction (Carter 1999, McCarthy 1998). Within Britain, interest was revitalized in 2004 by the publication of Carter's *Language and Creativity: The Art of Common Talk*, followed by, amongst others, Maybin and Swann's *The Art of English: Everyday Creativity* in 2006 and Aitchison's *The Word Weavers: Newshounds and Wordsmiths*, an exploration of journalistic creativity, in 2007. A repeated point is that linguistic creativity is a normal part of language behaviour:

So the main theme of the book is that creativity is an all-pervasive feature of everyday language. And, as I shall say more than once, linguistic creativity is not simply a property of exceptional people, but an exceptional property of all people. (Carter 2004: 13)

In this book, we suggest that the kinds of language creativity and artistry found in art and literature can also be found in the communication practices of everyday life ... We see language creativity as including both textual artistry and also the ways in which people use language creatively to construct identity and manage relationships with others ... (Maybin and Swann 2006: 1)

The kinds of lexical phenomena which they address can be exemplified by the following, taken from Carter (bold font added):

<S01>: He's at it again but he really wants you know just to sit down.

<S02>: Like they just talk about how they both feel.

<S01>: **Out of the frying pan into the deep freeze** this time.

(2004: 95)

A: He won't forget this time.

B: Brian, can you see **those pigs over my left shoulder moving slowly across the sky?**

(A and B both burst into laughter)

(2004: 23)

<S01>: There, that's solid now.

<S02>: I think I've made it **unsolid** ... sorry ... I've done it the wrong way round, have I?

(2004: 99)

As Carter discusses, the first involves wordplay, the manipulation of the idiom *out of the frying pan into the fire*, and metaphorical exploitation of *deep freeze*; the second alludes to the idiom *pigs might fly*, indicating disbelief; and the third non-standard affixation. Other kinds of creative usage include:

Rosamund, **I'm hotelled out**. I looked at c 100 hotels + I've got a short list of 4. You might like to look at their websites [etc.]. (written personal communication, December 2006)

Overheard also at Hay, aesthete and **uber-adapter** Andrew Davies wondering whether he really couldn't manage to "find a part in my next screenplay" for one of the nice Royal Ballet dancers performing so pleasingly ... (*The Guardian*, 1 June 2007)

where *be hotelled out* and *uber-adapter* are nonce-formations. Another kind relates to non-standard spellings, such as:

Heston's a renowned chef who specializes in **ker-azy** scientific cooking. He's best known for serving things like snail porridge and egg-and-bacon ice cream. (*The Guardian*, 20 October 2007)

Here *ker-azy* is used to suggest a lengthened and emphatic pronunciation of *crazy*. Similarly with *bares out* in the following (though most would consider this an error for *bears out*):

The [solar-powered] boat registers an electricity gain during a full moon ... It **bares out** research, that was greeted with scepticism some years ago, that claimed there was a minute, but measurable, increase in temperature during a full moon. (*The Guardian*, 11 December 2006)

Such cases are not institutionalized as lexical items: consequently, they could be considered outside the English lexicon, and beyond the scope of ordinary dictionaries which set out to describe that lexicon.¹ Yet they represent recurrent and productive patterns of usage: something that could be described as systematic creativity, and something that is very much part of the lexicon. By 'systematic creativity' I mean cases where individual words, phrases, and affixes are regularly used in creative ways to produce variations of meaning, including connotation and pragmatic effect. It is systematic in the same way that polysemy can be systematic (Apresjan 1974, Ostler and Atkins 1991, Nunberg and Zaenen 1992).² There is copious evidence for systematic creativity in everyday discourse, as Carter and others point out; there is also copious evidence in corpora, although corpora are generally used in lexicography to ascertain how language is 'centrally and typically' used (Hanks 1987: 124-5 and elsewhere) rather than eccentricities and exploitations, which are usually ignored.³

Thus systematic lexical creativity in English forms the topic of the present article. It looks in turn at four types of creativity: word meaning, affixation, idiom form, and respelling. For each type, it discusses corpus evidence, drawing on data from the 450-million word Bank of English (BoE);⁴ it then discusses

their treatment in dictionaries, specifically three recent British monolingual EFL/ESL dictionaries (see section 2.4.1). While the selection of examples to explore will necessarily be limited, I hope to demonstrate something of the extent of the phenomenon of systematic creativity, and the implications for lexicography, including the challenge to the normative function of dictionaries.

2. Words

2.1 Figurative uses

One obvious way in which words are exploited creatively is through figurative extension. Lexicographical choices are relatively straightforward: if the usage recurs sufficiently, it needs to be recorded, and the more established the usage, the more likely it is to be treated as an independent sense or subsense. For example,

It is a **perspective** which preserves a sense of the limitations of dependency theory and avoids the simplistic reductionism so common among the present-day **butterfly collectors** who abound in the social sciences and who **stroll** through history classifying types of dependency, modes of production, and laws of development, with the blissful illusion that their findings can remove from history all its ambiguities, conjectures and surprises ... (BoE: nonfiction)

Perspective 'viewpoint, stance' is historically a metaphor, but few dictionaries would label it as such, and many would prioritize this sense, the dominant usage, over any optical or architectural ones. In contrast, the metaphor *butterfly collectors* (not altogether transparent) is infrequent: though there is evidence in BoE of *butterfly collector/collecting/collection* being used by different writers to suggest an artificially abstract and somewhat opportunistic approach, the usage is not sufficiently institutionalized to be recorded as a sense, even in the unabridged *Oxford English Dictionary*.

The metaphorical use of *stroll* is transparent, and demonstrates a regular use of verbs/nouns relating to movement to describe intellectual and verbal behaviour (cf. the systematic conceptual metaphors described by Lakoff and Johnson 1980):

When you hear Balmain boy Alex Lloyd's debut album, *Black The Sun*, you **take a stroll through** rock music's past, present and future.

Burke invites you on an odyssey that begins in the present day and **travels through** history to bring to life the remarkable nature of inter-connected events and discoveries.

I was rather surprised by how little encouragement the students were given to **wander off down the byways** of the subject.

Yet BoE contains only isolated instances of noun/verb *stroll* with such a meaning, and so this too is probably one to be considered as an ad hoc usage,

and ignored — the only really recurrent metaphorical uses of *stroll* in BoE relate to easy victories:

And — four days after the UEFA Cup win in Germany — they **strolled to victory** over Airdrie to book a place in the League Cup final.

Dealing with cases like these (or the metaphorical use of *deep freeze* in the very first example in Section 1) is part of normal lexicographical practice with respect to the analysis and representation of polysemous words. Figurativeness is part of systematic creativity, but it is not really this kind of creativity which concerns me here.

2.2 The construction of figurativeness

A more intriguing case, though, is that of words which seem to create figurativeness in their co-texts. For example, a minor subclass of English adjectives comprises items such as *archetypal*, *proverbial*, *quintessential*, *ubiquitous*, *veritable*, and so on: they are mainly formal and/or Latinate, and interesting because of their functions in discourse, rather than their literal meanings. Only rarely does *proverbial* mean 'relating to proverbs'; far more often it is used to mark the use of a particular linguistic item — occasionally a proverb, more often an idiom or part idiom, metaphor or connotative item, and sometimes substituting for a word euphemistically. It is a preemptive device, distancing writers/speakers from their choice of lexis. The following BoE examples are typical:

to squeeze through the eye of the proverbial needle). <p> But, alas, it's not fine and that Tailhook is just the proverbial few bad apples, or that there-me election, we Tories were given the proverbial boot, and the more soberly dresse scarcely large enough to swing the proverbial cat. Instead, builders are creati that's for sure. Poor as the proverbial church mice, we are now. In fact has now taken to showing like the proverbial duck to water." <p> Roy Creber's will not sell. What will sell like proverbial hot cakes is his story about Ferg had not ruled her family with the proverbial rod of iron, but with an influec omparison had been financed on the proverbial shoe-string." P. G. WODEHOUSE aut but eventually caught on like the proverbial wildfire; Bagatelle acquired a lo

turing qualities. Here was the proverbial able woman behind a gifted man. things and study things. It's the proverbial aim, aim, aim, aim. And never pul can feel uncomfortably like the proverbial bacillus invading a weakened have any content). It's geeky (the proverbial boys in their bedrooms). Digital is a hit. Our visit was on the proverbial dark and stormy night. To get to ental and spiritual satiety on the proverbial desert island. <p> Not content wi she said. Doran looked like the proverbial old friend who arrives at the repeated as drearily as the proverbial old broken gramophone record. She had a constitution like the proverbial ostrich. She never knew what it us. We didn't shout or throw the proverbial plates at each other. Why should sustained from falling down the proverbial slippery staircase. One common y that was made in private, in the proverbial smoke-filled rooms, and sealed by ket. It has continued to climb the proverbial wall of worry after the briefest scratches as it advances up the proverbial ladder of success. At long last spoiled rotten as a child - the proverbial 'unsatisfactory eldest son' -

joy - or they can be a pain in the proverbial. The same can be said of partners and remain mad about the son-of-a-proverbial. They say that they know they sho

It is the second grouping which represents systematic creativity, where *proverbial* signals something which is meant to be interpreted in a figurative or symbolic way.

Similarly, *ubiquitous* could be glossed as 'occurring everywhere', but BoE evidence suggests that it is normally hyperbolic. In particular, *the ubiquitous* seems to indicate that the reference of the following noun group is symbolic, connotative, and something of a cliché (*the ubiquitous club sandwich, grey suit, droopy rucksack, sneakers*):

owed a great deal to Macintyre's "ubiquitous and enthusiastic support". He was to come in on the act. Clays are ubiquitous at the Earth's surface. Under the of a superior arbiter, such as ubiquitous British ministers, local hardening delights rather than the ubiquitous club sandwich. Appetisers include girl, one shoulder bearing the ubiquitous droopy rucksack and the other a a million square metres of the ubiquitous floor covering) was still selling <p> A host of tiny villages, ubiquitous flower pots, whitewash and grey get a degree in law, acquire the ubiquitous grey suit (only 12% of chief and retro clothes in contrast to ubiquitous hippie gear, quickly became the is in many fats and protein is ubiquitous in food (meat, fish, eggs, nuoc mam, the fermented fish sauce ubiquitous in Vietnamese cuisine, wafted up ever popular Border Collie and the ubiquitous Jack Russell Terrier. <p> Dogs and pork), shin of beef, the ubiquitous poulet Brittany produces a third hospital or clinic. Not even the ubiquitous radiation hazard signs or the DAMN QUEER </h> <p> Venue: The ubiquitous "Secret London Location". Makes a is the one thing, along with the ubiquitous sneakers, that even the most cash or less. And then this: rot and ubiquitous tropic heat and atmospheric a world different from our own ubiquitous tropics. But no more! And, subtle alternatives to the once-ubiquitous vindaloo and Madras. 'We've come Our distribution networks are ubiquitous worldwide. We have the leading

Veritable has no real denotative meaning at all, but functions as an emphaser:⁵

Society and its ilk, Wallace was a veritable angel of deliverance from the 1900. Not to be missed ... No. A "veritable apotheosis", according to the like, exists. <p> Omnibus are a veritable conveyor belt for this sort of performances. <p> London saw a veritable flood of Merlin dramas at this week's model is old news. There's a veritable gaggle of new machines heading Edwards and Benson Perkins, a veritable galaxy of British Methodists. budget film, The Eagle thought. A veritable hail of bullets was coming the blind dates as if epiphanies, veritable life-changing afternoons. The The Symphony, for example, is a veritable masterpiece of compression. for catapulting her into this veritable mayhem of wealth, luxury and remains juicy. <p> The veal chops – veritable menu workhorses – are compatible approach with care and caution the veritable minefield of the ecos, latter-day the wettest place on earth. It is a veritable odyssey, full of magic and for the snap? Our walk becomes a veritable royal walkabout and I experience London W8 (071-938 2711) is a veritable souk of ceramics, table lamps, contribution. This book is a veritable storehouse of information for the Well, the last missive is the veritable straw. How can we, as governors, Noam Chomsky has issued a veritable torrent of passionately written, I regarded the local library as a veritable treasure trove. There was no a book.' Some hope. Weekends are a veritable whirl of activity. After a

The typical metaphorical status of the following noun is further demonstrated by a list of the most significant items found in BoE after *veritable*: in descending order, using mutual information statistics, these are *cornucopia, smorgasbord, potpourri, fountain, menagerie, Aladdin's cave, hotbed, goldmine, feast, minefield, encyclopedia, inferno, banquet, blizzard, torrent, plethora, treasure (trove), avalanche*.

Many of these collocates — and others — refer to complex or multi-faceted entities.

These words are not intrinsically creative, but they all signal some kind of exploitation of the co-text: in particular that the following item is not to be taken literally, but understood in terms of its connotations and implications. This in itself is a creative use of referential meaning.

2.3 'I'm hotelled out'

While English phrasal verbs are often regarded as idiomatic and idiosyncratic combinations, there are consistent patterns of usage, with consistent meanings of particles. *Out*, for example, occurs in combinations relating to eradication and disappearance (*airbrush out, die out, phase out, weed out, wipe out*); depletion and exhaustion (*run out, sell out, tire out*); completeness (*carry out, flesh out, sort out, spread out*); and allocation (*mete out, parcel out, share out*). Such meanings may be explained in dictionary entries for particles, and their systematic metaphoricality is discussed by, among others, Kövecses and Szabó (1996) and Lindstromberg (1997). *Be hotelled out* fits best with combinations referring to exhaustion (*be burned/played/tired/worn out*), and closely corresponds to informal *be partied/shopped out*, which recur in BoE, and other such combinations which do not.

If you are still **all partied out** after the New Year, pick up a pick-me-up.

... the Priory group of hospitals — famous for treating the addictions of **partied-out** celebrities.

Some feel **shopped out** after a grand fling with consumerism during the mid 80s.

We decided we were **all business-conferenced out** and we wanted to help young entrepreneurs find balance in their lives.

All **Buena Vista'd out**? Cachaito Lopez will get you back into the swing, says Nigel Williamson.

This suggests a specific creative phrasal verb frame, *be VERBed out*, where VERB is supplied by a word that is predominantly used as a noun, with the whole referring to a state of having had too much of the thing denoted by the noun.

2.4 Dictionary treatment

2.4.1 The dictionaries

To explore how current dictionaries deal with such phenomena, discussion here focuses on treatment in three British monolingual dictionaries for advanced learners of English as a foreign or second language: the *Longman Dic-*

tionary of Contemporary English (4th edition, 2003, henceforth LDOCE4), the *Oxford Advanced Learner's Dictionary of Current English* (7th edition, 2005, henceforth OALD7), and the *Macmillan English Dictionary for Advanced Learners* (2nd edition, 2007, henceforth MED2). All three dictionaries aim primarily to cover mainstream vocabulary, concentrating on central usage, and their pedagogical function means that they have to be normative: to expect them to deal with creative usage might therefore seem unreasonable. Yet these dictionaries, perhaps more than larger dictionaries for first-language speakers, are very responsive to language change and they have developed a variety of flexible techniques for describing different linguistic phenomena. The three selected dictionaries, the three market leaders, have undergone major revisions in recent years, taking corpus evidence into account: MED was written from the outset with corpus data. Moreover, all need to keep their target users (students in their late teens or early twenties) clearly in mind, along with the kinds of English lexis and globalized discourse to which those users are exposed, including electronic media where creative phenomena abound. In many respects, therefore, they represent a cutting edge of lexicography, driven on by a changing market and the need to show clearly how lexis actually works in current English around the world.

2.4.2 Some entries

To begin with *proverbial*, all three dictionaries treat it in terms of its usage in collocation with fixed expressions:

LDOCE4: **the proverbial sth** used when you describe something using part of a well-known expression: *The store had everything including the proverbial kitchen sink.*

OALD7: used to show that you are referring to a particular proverb or well-known phrase: *Let's not count our proverbial chickens.*

MED2: used when you are describing something with an expression from a proverb: *He turned up just like the proverbial bad penny.*

However, these explanations do not account for collocations with other figurative items, nor the way in which *proverbial* signals that the following noun is to be interpreted metaphorically or in terms of connotations.

Ubiquitous and *veritable* receive contrasting treatments. The first is labelled 'formal', and given definitions which cover its literal meaning; its weakened sense is conveyed through *seeming* and *humorous(ly)* in LDOCE4 and OALD7, and in all three through the implied hyperbole of examples (where the choices of noun collocates to mention may not be as meaningful to dictionary users as lexicographers).

LDOCE4: *formal* seeming to be everywhere — sometimes used humorously: *Coffee shops are ubiquitous these days. | a French film, starring the ubiquitous Gérard Depardieu.*

OALD7: (*formal or humorous*) seeming to be everywhere or in several places at the same time; very common: *the ubiquitous bicycles of university towns* ◇ *the ubiquitous movie star, Tom Hanks*.

MED2: *formal* present everywhere: *Their HQ comprises miles of corridors with their ubiquitous coffee machines*.

The pragmatics of *veritable*, however, are shown explicitly in definitions, with MED2 also offering a partial paraphrase *real*; LDOCE4 gives *real* as a 'synonym', OALD7 gives *positive*. All label *veritable* as formal, with OALD7 adding 'or humorous'.

LDOCE4: *formal* a word used to emphasize a description of something: = **real** *The area is a veritable paradise for those who love walking and swimming*.

OALD7: (*formal or humorous*) a word used to emphasize that sb/sth can be compared to sb/sth else that is more exciting, more impressive, etc. SYN POSITIVE: *The meal that followed was a veritable banquet*.

MED2: *formal* **real**: used for emphasizing what you are saying: *a veritable feast* ♦ *a veritable army of helpers*

In each case, examples show metaphorical and hyperbolic noun collocates for *veritable*, but nothing more is said or shown about the typicality of this kind of collocation. (In contrast, the *New Oxford Dictionary of English* (1998), a large monolingual dictionary for native speakers, says of *veritable* 'used as an intensifier, often to qualify a metaphor: *the early 1970s witnessed a veritable price explosion*'. However, it does not treat the other items and usages examined here in any more depth.)

The dictionaries could not be expected to cover *be hotelled out*, or really the relevant use of *out* — though MED2's study page feature on word formation comments on phrasal verb particles having recurrent meanings. Particle meanings are discussed more fully in some EFL/ESL dictionaries of phrasal verbs. For example, the *Collins Cobuild Dictionary of Phrasal Verbs* (1989) had a separate appendix on particles, and *be hotelled out* seems to fit with a category 'ending of activities', where the discursive explanation includes

... Some combinations indicate that an activity cannot continue any longer, because all the things or people involved have been used up or destroyed ... These combinations are often passive. For example, if you **are played out**, you are exhausted and feel unable to do anything.

Macmillan Phrasal Verbs Plus (2005) has a series of sophisticated entries for particles in the main alphabetical sequence of the dictionary: here graphics are used extensively to show relatedness of meanings. The closest senses to that in *be hotelled out* are those represented under *out* in relation to 'removing':

have nothing left, be finished (run out, sell out)

stop existing, end, disappear (burn out, fade out, go out)

But far more significant is the dictionary's special feature 'New' Phrasal Verbs in its pedagogical mid-matter. It explicitly discusses creativity and this specific use of *out*:

In 2003, the celebrity cook Delia Smith was quoted as saying that she was 'all *reciped out*' when she announced her intention to retire from TV cookery programmes ...

... writers and speakers often use words in creative ways, but dictionaries do not generally describe these individual creative acts. Nevertheless, we understand what Delia means, because the expression she has created follows a pattern that we are already familiar with ...

A later part of this feature talks about combinations such as *be partied out*. This feature provides one of the strongest acknowledgements of creativity in current learners' dictionaries, both general and specialized; but it is not part of the main alphabetical text, and the entry for *out* itself does not cross-refer to the discussion.

To summarize this section, it could be said that the dictionary entries discussed in this section are reasonable attempts to cover usage; the issue is that more could be said, and the extra might help explain a systematicity which underlies these kinds of usage.

3. Word formation

3.1 Affixation

In an example given in Section 1, the designation *uber-adapter* was applied to the British screenwriter Andrew Davies, who is noted for his imaginative and elegant adaptations for television of literary novels, including classics such as *Pride and Prejudice* and *Bleak House*. *Uber-* is an odd prefix in English. German in origin, it echoes *Übermensch*, which is now considered unacceptable because of its eugenicist and racist historical associations, yet these associations do not cross over into English, where it occupies something of the semantic space of *super-*, *mega-*, *hyper-*, as a slightly ironic, currently fashionable, indicator of notoriety, superiority, or domination. In terms of morphology, *uber-* scarcely occurs at all in institutionalized words — other, perhaps, than *uber-babe* or *uber-cool* — but is mainly found in a considerable number of ad hoc formations. Those attested in BoE fall into sets: relating to young people (*uber-babe*, *uber-brat*, *uber-geek*, *uber-nerd*, *uber-teens*); the entertainment industries (*uber-director*, *uber-entertainer*, *uber-guitarist*, *uber-rockster*, *uber-star*); other careers or origins (*uber-capitalists*, *uber-careerist*, *uber-chef*, *uber-lawyer*, *uber-Dubliner*); and, in adjectives, to fashionable status (*uber-chic*, *uber-cool*, *uber-glam*, *uber-groovy*, *uber-hip*). Thus it is used creatively, to label and emphasize, and also to mock, partly by distancing the writer/speaker from both label and labelled.

The suffix *-esque* has a much longer history in English, occurring after nouns or names in well-established adjectives such as *picturesque* and also

statuesque, rubenesque, junoesque, referring to (mainly female) body types. It is used productively after the names of writers, artists, entertainers, politicians and so on, to form adjectives which indicate or allude to qualities typical of the person mentioned. Better-attested formations, sometimes hyphenated, in BoE include *Kafkaesque, Clintonsque, Disneyesque, Beatlesque, Dylanesque, Reaganesque, Abbaesque, Kiplingesque, Chandleresque, Daliesque, Pinteresque, Stalinesque, Capraesque, Chaplinesque, Garboesque*, but the creative potential is perhaps best demonstrated in context, in this case with a sample of hyphenated forms where names begin with W:

narrative involving a group of Waco-esquereligious oddballs on a remote
 s a piano-accompanied, Tom Waits-esqueballad and an acoustic guitar-led,
 the bloke obscured by a Rick Wakeman-esquewall of keyboards starts to look
 be dressed up to look like a Waltons-esquelifestyle choice. <hl> Porsche
 to reinstate automobiles as Warhol-esquewhite-trash pop icons, with ram-
 misterieuses, from the Watteau-esquefrivolity of L'Arlequine to the
 <p> Petit sighs, the pursuit of a Wax-esquesoundbite, not to mention a
 make her gnash her teeth in a Wharton-esquedisplay of pique. 'People stop
 hours trying to make my own Wheatley-esquebooks. <p> Recently, I have enjoyed
 though some of the Gilbert White-esquecharm was lost in the process.
 Coogan's mind, this Richard Whiteley-esqueceleb is almost a living entity. He
 verse along with the more Whitman-esquestuff (why do Americans like Walt
 songs, plays most of the chopping Who-esqueriffs that punctuate them, and has
 3 girl in mufti, the other Widdecombe-esquein vintage M&S. Kidman clasped
 to unveil the appalling Kim Wilde-esqueballad Making It Happen". While
 of pop with a hint of Robbie Williams-esquecrossover chucked in. And why not?
 exiting the yard at a Matt Williams-esquerate, the District 52 Little League
 appearing here on the Steve Wonder-esqueDon't Let It Go To Your Head"),
 bottle with an Alice In Wonderland-esquelabel: 'Rub this into your horny
 could result in a Brave New World-esqueworkplace, where all employees are
 adopt during sex: very much the Wyman-esquebass players of the business, while

These mainly occur in journalistic components of BoE, and include references to places (Waco), books (*Alice in Wonderland, Brave New World*), and entertainers, TV personalities, and celebrities: all symbolic of current culture, though references and relatedness are essentially vague. Very few of these ad hoc or creative formations evaluate positively; most are negative, and imply 'less good than' or 'imitative'.⁶ This contrasts with the better-attested formations, which mainly have more specific meanings and often neutral or positive evaluations.

3.2 Analogy and blending

Other new words are formed through analogy or blending, by being modelled on an existing word rather than by combination of established morphemic elements. One such case involves *literati*, a 17th-century borrowing into English from Latin, and used to refer to highly-educated and cultured people, sometimes with an implication of elitism:

of 'fashion and the longings of literati', and there is no need to say
 with a large number of vieux Soholiterati, the ones that go in for very long
 face and announced to the assembledliterati that it was good to meet a real
 83) and the available Concordliterati and transcendentalists; more often
 Net Book Agreement; unlike other literati, the average Irish writer tends to

This has generated a blend *glitterati*, referring to celebrities in the entertainment and other industries:

dancing and an entranced crowd of glitterati in evening dress who watched the form outside, the fashion industry glitterati who drop in for bangers and mash-unsophisticated, not Washington glitterati, they were all young and came list of freebies, lots of local glitterati and party faithful. <p> All the Got a Woman and All Shook Up. The glitterati were jetted in – as was Sam

BoE has 185 tokens of *literati*, 196 of *glitterati*. It also provides evidence of many further formations with an ending *-erati* and indicating cliques of people who are characterized by a particular lifestyle, interest, or trait. There are 8 tokens for *chatterati* (people who regularly appear as pundits on television and in the media generally, sometimes referred to as 'the chattering classes'): for example,

the anti-science views of the chatterati. Science actually works. And it in the right place. To the media chatterati, to be 'vulgar' or Northern has And I think there's a high-level chatterati perception, which is, I guess,

Of other formations in BoE, *rockerati* occurs 3 times, the rest just once or twice:⁷

all 071-437 2668 #pound; Mingle with the arterati at the Design Museum's superb radio interviews with the welfare benefiterati, began during the summer of not have much influence with the committerati of New Britain. <hl> Britain latest smash among the Hollywood fitterati? Believe it or not, the answer is sense is small solace, though, for flitterati anxious to look good on the off g chi-hiked by the White Hart Lane lagerati but, should Tottenham again Cooper Barr, leading lights of the mcglitterati and owners of The Apartment rt of 4AD's '13-Year Itch" and the poserati are out in force. Just to which seem to be popular with the rockerati. There is an interesting tension, a training track for the capital's scooterati. At the station, board a ladettes of the newly-enfranchised soccerati. Scousers, naturally, hate every by the literati, glitterati and twitterati of the period. Now, juxtaposed Garrison Keillor. Clearly, if the witerati is about to conquer the world of

While *literati* can be used neutrally, almost all other formations are used ironically and deprecatingly.⁸

3.3 Dictionary treatment

The learners' dictionaries treat word formation in two ways: through individual entries for the most productive affixes and through special features, which explain principles and systems, and/or list, gloss, and exemplify affixes. *Uber-* is explained as follows:

LDOCE4: *informal* better, larger, or greater; = **super**: *uberbabe* Pamela Lee | *I want to do something uber-cool with my webpage.*

OALD7: (*informal*) (in nouns and adjectives) of the greatest or best kind; to a very large degree: *His girlfriend was a real uber-babe, with long blonde hair and a big smile.*
◇ *The movie stars the uber-cool Jean Reno.*

MED2: *often humorous* **1** very: used with adjectives: *My new sneakers are uber-cool.* **2** total, or complete: used with nouns: *This is the web page of an uber-nerd.* ♦ *jet-setting uber-babe Mischa Barton.*

These entries are not inaccurate, but nor do they satisfactorily explain combinatorial patterns and the manneredness or negativity of less recurrent forms. Entries for *-esque* are as follows:

LDOCE4: **1** in the manner or style of a particular person, group, or place: *Kafkaesque* (= in the style of the writer Franz Kafka) **2** having a particular quality: *picturesque* (= pleasant to look at)

OALD7: in the style of: *statuesque* ◇ *Kafkaesque*

MED2: **1** used with the names of some people to make adjectives describing things relating to that person or their work: *Kafkaesque bureaucracy* **2** used with some nouns to make adjectives describing things that have a particular quality: *a picturesque landscape*

The most striking point here is that *Kafkaesque* is used to illustrate the formation, yet this is hardly compositional: it has a particular meaning 'nightmarish, bewildering', alluding to situations described in Kafka's work, rather than simply describing his work or style — and in fact both OALD7 and MED2 have separate entries which well describe exactly this sense.

Literati is a more complex case: it has become a model, but can scarcely be decomposed into a suffix *-erati*. All three dictionaries treat *literati* and *glitterati*; MED2 has *chatterati*, though with a different meaning from Indian English. Nothing more is said, and perhaps all that could be added is some kind of cross-referencing between the forms, or a comment that analogous hapax forms occur.

4. Idioms

4.1 Variability and instability

Idiom is an ambiguous term, but is most narrowly applied to a subclass of fixed expressions with non-compositional meanings: metaphorical phrases such as *bury the hatchet*, *spill the beans*, *through thick and thin*. Ad hoc exploitations are regularly found — for example, BoE has *spill the enchiladas/flageolets/haricots* — and corpus research in general has provided extensive evidence of idiom instability and variation: see, for example, Moon 1998: 120ff.

To begin with the idiom exploited in the example 'Out of the frying pan into the deep freeze ...' (Carter 2004: 95), typical realizations in BoE are:

go out of the frying pan (and) into the fire
jump from the frying pan into the fire

it's out of the frying pan into the fire
out of the frying pan ... [with ellipsis]

Exploitations include, with the first paralleling Carter's example:

Escaping the heat of Calcutta, it's **out of the frying pan into the freezer**.

In the last few months, however, these adventurers have discovered that they have **leapt from the frying pan into the flames of the deepest recession** to hit America since the 1930s.

I really liked Kenneth. But it was a case of **out of the frying pan and into the volcano**. Sometimes he would disappear and phone to say he was on his way home then take six hours to get there.

Greece [sc. football team] proceeded **from the fire into the frying pan** on Saturday ...

In the case of *pigs might fly*, typical variations are *pigs may/will/can't fly*, and BoE also provides many examples such as:

Do you believe the British press would behave honestly and calmly, without malice or bias? Look out of the window. **Is that a pig flying by?**

But **until perfect pigs are spotted flying in the sky**, they would be wise to stop promising things they cannot deliver.

They **were looking for flying pigs** when the West Indies set off in pursuit of a victory target of 364. What they got was the sight of chickens coming home to roost as they collapsed pitifully to 51 all out, their lowest Test score.

The forms in which traditional proverbs are found can also vary greatly: many are less often found as independent clauses (*a drowning man will clutch at a straw; it's the last (straw) that breaks the camel's back*) than as verb + complementation idioms (*clutch at straws*) or noun compounds (*the last straw*). For example, BoE has examples of the proverb *every cloud has a silver lining* but many more of *silver lining* on its own (around 60% of all tokens); further examples allude in some way to clouds, which represent metaphorically a disadvantageous situation, and some even reverse the form of the proverb:

There's no point in being depressed about it because **every cloud has a silver lining**.

Indeed, yesterday was one of those occasions when policy-makers seemed intent on finding **a silver lining**.

The only silver lining in the clouds is the unofficial opposition, the unsung heroes and heroines, who despite the massive and near-absolute power of the state, have never given up the struggle ...

This is the sparkle of gold in the dust of unemployment, **the silver lining in the cloud of apathy**, the light at the end of the tunnel of hopelessness.

On the surface then, Ronnie O'Brien's move to Italy was **a silver lining** at the end of this **cloudy** spell.

The bay area may be the one major city that's had a good '92, but it's a case of **every silver lining having a larger black cloud**. The Athletics won their division but then lost the playoffs because their hero, the unloseable reliever, Dennis Eckersley, lost when it counted.

Perhaps the most creative idiom phenomenon is the frame: cases where there are no fixed lexical words at all but an underlying structure which signals and creates the idiomatic meaning. One such can be glossed as 'unintelligent, foolish', and its best-known realizations include:

two sandwiches short of a picnic
several cards short of a full deck
a brick short of a full load
a few beers short of a six-pack

BoE has, however, many further realizations, which are intended to amuse and typically relate cohesively to the context, or represent metonymically the person being described or their activities. Creativity is thus the norm:

a few gallons shy of a full tank
several leaves short of a complete lettuce
one banana short of a bunch
a cake short of a party
a couple of sprouts short of a Sunday luncheon
several petits fours short of an Ascot hamper
at least a couple of desks short of a classroom
several test tubes short of a laboratory
a couple of stethoscopes short of a full set of surgical instruments
only a couple of custard pies short of a circus act
two thumb-screws short of a torture chamber
a few slates short of a roof
a bishop short of a chess set
a watercolour short of a palette
two sheep short of a flock
three diamond clusters short of a tiara

4.2 Dictionary treatment

General dictionaries tend not to allocate much space to metaphorical idioms such as these, so only the best established of variations are shown, examples are sparse, and information is aimed primarily at decoding. Explanations of *out of the frying pan ...* in the three learners' dictionaries are broadly similar; none give examples, nor show collocating verbs:

LDOCE4: **out of the frying pan and into the fire** to go from a bad situation to one that is even worse

OALD7: **out of the frying pan into the fire** (*saying*) from a bad situation to one that is worse

MED2: **out of the frying pan (and) into the fire** used for saying that someone who was in a bad situation is now in a worse situation

All three dictionaries cover *pigs might fly* by explaining its pragmatics, and, in the case of OALD7, the American variant *when pigs fly*:

LDOCE4: *spoken* used to say that you do not think something will happen: 'Someone might have handed in your pass.' 'Yes, and pigs might fly.'

OALD7: (*ironic, saying*) used to show that you do not believe sth will ever happen. 'With a bit of luck, we might be finished by the end of the year.' 'Yes, and pigs might fly!'

MED2: *British spoken* used for saying that something is completely impossible

All three dictionaries treat *every cloud has a silver lining* at the headword *cloud*,⁹ but not the elided form *silver lining*.¹⁰

Two sandwiches short of a picnic (*etc.*) is covered in two of the learners' dictionaries, under *short*:

LDOCE4: **be one ... short of a ...** *spoken* used humorously to say that someone is a little crazy or stupid: *Lady, are you a few aces short of a deck? | He's one sandwich short of a picnic.*

OALD7: **a brick short of a load, two sandwiches short of a picnic, etc.** (*informal*) (of a person) stupid; not very intelligent.

LDOCE's citation form is perhaps harder to interpret than OALD's, but the definition covers pragmatics more successfully.

Variability is a significant aspect of idiom behaviour, and some attempt to acknowledge this is made in the special features on idioms in LDOCE4 and OALD7 (though not MED2). LDOCE4 has:

Idioms are normally used in a very fixed and limited way. Often, though, English speakers play with idioms by making up their own versions of an existing idiom.

It illustrates this with the pattern *it drives me crazy/nuts/mad/up the wall/bananas*. OALD7 has:

In some idioms, many alternatives are possible. In the expression **disappear into thin air**, you could replace **disappear** with **vanish, melt** or **evaporate**. In the dictionary this is shown as **disappear, vanish, etc. into thin air**, showing that you can use other words with a similar meaning to disappear in the idiom ...

Both dictionaries here refer to the more tractable side of idiom variation, and it is perhaps a pity that space constraints did not permit them to add something about more extreme and creative forms.

Specialist dictionaries of idioms have more scope, and, for example, particular attention is paid to variation in the *Oxford Dictionary of English Idioms* (1993, henceforth ODEI),¹¹ which has a special index of variations, and the *Collins Cobuild Dictionary of Idioms* (1995, henceforth CDI). To demonstrate something of the possibilities: ODEI's entry for *pigs might fly* has a slightly overliteral definition ('wonderful, and seemingly impossible, things may happen'), but its three examples include a gross variation and pragmatics-oriented note, while CDI adds a note on variability:

ODEI: ... (NONCE) *If each individual were morally regenerated ..., the world could become a better place. Equally, if pigs had wings, they might be taught to fly.* □ variant 'if pigs had wings, they might be taught to fly'; implies speaker does not believe that things referred to will happen.

CDI: People often vary this expression, for example by saying they saw **a pig flying by**. 'Maybe one day we'll be seen as entertaining.' 'Oh look, I just saw a pig fly by my window.'

ODEI treats *every cloud has a silver lining* as a headphrase form, but examples show variation including the reversal, and a note on variability and pragmatics (CDI's long entry deals with truncation, reversal, and exploitations):

... *Patrick decided things were far from bad just now. Even an alcoholic cloud may turn out to have a silver lining if taken in time.* □ (NONCE) *He turns out to be considerably less pessimistic than most writers whom we think of as 'serious', for whom every silver lining was a cloud* □ often adapted; often said to encourage oneself or sb else.

The 'two sandwiches' frame is not included in ODEI, but CDI comments explicitly on its variability and creativity, and its examples, omitted here, cover two 'canonical' variations and two creative ones:

Short of is used in expressions such as 'one sandwich short of a picnic' or 'several cards short of a full deck' to indicate in a humorous way that you think someone is very stupid or behaving very strangely ... This expression is used very creatively, and people often use it simply for the humorous effect of a new and amusing variation.

5. Respellings

5.1 Changing forms, changing usage

My final examples of linguistic creativity relate to spelling. Ad hoc creativity regularly occurs, for example in newspaper headlines and in electronic personal correspondence, whether private or semi-public, but corpus data is not

really the best source of data to investigate it. However, some respelled forms such as *ker-azy* do recur in BoE:

... find yourself hailed as Harbingers of ker-azy new scene.
His guests include the Doug Anthony All-Stars, ker-razee comedians.
... ker-razy beer boy singer Jerry A ...

More relevant for lexicography are cases where respellings become institutionalized, developing special meanings and usages of their own. *Lite*, originally an informal American respelling of *light*, began to be used as an indicator that food had less fat or calorific content, or drinks had less alcohol, than standard:

The meaning of 'natural' has been subverted by being coupled with 'diet' food. Diet food, also called '**lite**' food, accounts for a major proportion of the food industry's growth.

The Marks & Spencer **Lite** yogurt range is an unbeatable combination of fruit and wholesome yogurt.

It is now regularly used, postnominally, to deprecate something as an oversimplified or weakened form of an original, more challenging or demanding, activity, subject, and so on:

The result is that he serves us up with **history lite**, something much less appetising than the real thing.

Martin Amis ... objects that the biography and the interview are a sort of **Literature Lite** for those too lazy to take the real thing: 'The truth is that we are more interested in writers than we are in writing.'

Environmental groups criticised the 'watered down' deal. Greenpeace described it as '**Kyoto-Lite**'.

Pastels, a sort of '**colour lite**' for those too scared of the real thing, run the real risk of looking twee and faint-hearted.

Similarly, in British English the respelling *yoof*¹² is used to refer, often disparagingly, to youth culture and language, while *girrl* (with varying numbers of *r* and *l*) denotes a spirited and confident young woman:

post-war explosion of materialism-madyoof culture in which Feld's Bolan-ness
Somewhere between the horror of "yoof TV" and the shambles of trendy late
over bumpy acne, and painful 1980syoof -speak. Young people these days

whisky on me. <h> Here come the girrls </h> <p> If two can make a trend, I
while kissing Zeta Jones "a great girrrl , very feisty, very punctual" was 'a
and declared: 'You can't have that girrrllplaying Thomas a Becket #pound; Why

These are spelling innovations which, arguably, fill semantic gaps.

The case of *it bares out research* is perhaps different. While it would normally be considered an error (though this form *could* be interpreted as realizing the same metaphor as in *lay bare*, 'make evident, reveal'), it is representative of

other cases where erroneous spellings co-exist with standard ones. Many of these involve homophonous or near-homophonous words, as in the following, where the canonical or original form is given first:

give something free rein, give something free reign
 give full rein, give full reign
 off one's own bat, off one's own back
 whet sb's appetite, wet sb's appetite

Evidence from corpora, web data, and elsewhere suggests that the 'non-standard' forms are now in free variation with their originals, and may well supplant them. As with *bear/bare out*, respelled/mispelled forms are rationalizable, for example, through different metaphors: *off one's own bat* is a cricketing metaphor, *off one's own back* could suggest a metaphor to do with getting up and becoming active; *whet someone's appetite* involves a metaphor to do with sharpening,¹³ *wet someone's appetite* suggests a metaphor to do with providing an appetizer or aperitif.¹⁴

5.2 Dictionary treatment

Only the best-established non-standard spellings feature in learners' dictionaries. All three have *lite* in the context of low-calorie/-fat foods and low-alcohol drinks, but only OALD7 has a second sense:

2 (used after a noun) (*disapproving*) used to say that a thing is similar to sth else but lacks many of its serious or important qualities: *I would describe this movie as 'Hitchcock lite'.*

All three deal with *yoof*, though none (unsurprisingly) with forms of *girrl*:

LDOCE4: *adj* [only before noun] *BrE* relating to or intended for young people — used humorously: *a yoof magazine*
n [U] *BrE* young people, considered as a group — used humorously: *British white yoof*

OALD7: (*BrE, informal, humorous*) a non-standard spelling of 'youth', used to refer to young people as a group, especially as the group that particular types of entertainment, magazines, etc. are designed for ► **yoof** *adj.* [only before noun]

MED2: *British very informal humorous* young people. This word is used especially on television, in the newspapers etc, as a humorous way of spelling the word 'youth': *today's yoof* ♦ *the new face on yoof TV*

The three attribute *yoof* to different registers; the label *humorous* is perhaps intended to subsume the slight disparaging quality of the use of *yoof*.

None of the three dictionaries covers erroneous or changing spellings of the idioms mentioned. At a more general level, though, MED2 deals with creative respelling in its study page feature on word formation:

... Sometimes they develop different meanings from the original versions: for example, the word **wannabe** is a respelling of the phrase *want to be*, but it is used as a noun meaning someone who wants to be famous or successful.

6. Implications

In the preceding sections, I have discussed a range of lexical phenomena which demonstrate something of the systematic creativity of language in use. The cases selected were simply some which I had noticed and investigated: not, of course, a scientific approach. Yet I would argue that they are nonetheless representative, since for each type, whether loose metaphorical use, pragmatic signal, frame, affix, idiom, or respelled item, many more examples could be found with similar productive patterns of creative usage: hence their systematicity. My choice of dictionaries has also been deliberately limited, and perhaps others succeed in areas where these, arguably, do not — though in fact the three are all very successful dictionaries, written with great skill and thought, and taking into account a variety of evidence sources; moreover, each one deals well with individual instances of a creative phenomenon. I must emphasize, too, that lexicographers are certainly not unaware of creative phenomena, since these uses are precisely the ones which they separate from 'central and typical' usage in the evidence. The question is, however, whether such uses need be entirely discarded in the way that they so often are.

A whole array of creative linguistic phenomena are readily findable in corpora, as search engines permit. These are not just oddities or ad hoc neologisms, but realizations of systematic creativity in language. Smaller dictionaries are necessarily limited in what they can do or show or explain, but by omitting such phenomena, or by treating them 'lightly', they neglect an important aspect of language. This has consequences for dictionary users with respect to decoding, because of the connotations and pragmatics: there are also consequences for encoding, since comfort and confidence with a language comprises more than just fluency and grammatical competence. After all, language learners and non-native speakers, as native speakers, are not oblivious to creativity, puns, and metaphoricity. Lexicographers may therefore be doing their users a disservice: in presuming that users do not want or need further comment or coverage, in neglecting the evidence of corpora, and, not least, in passing up an opportunity to communicate the sheer fun of a language as well as ways in which usage is practised or changes.¹⁵ If dictionaries do not discuss these creative phenomena, it is difficult to see how or where users can find reliable information about them, and thus find out about the meanings of the language and discourses with which they are engaging. This, it seems to me, is frustrating for everyone, not least lexicographers themselves.

Notes

1. In contrast to 'subversive' dictionaries, such as those of slang or neologisms, though these are generally more interested in actual formations than patterns of formation.

2. For example, where words change word class (noun = form of transport → verb = be conveyed by that transport: *bike, ship, motor*, etc.) or nouns change countability (uncountable noun = substance → countable noun = types or portions of that substance: *cheese, wine, bread*, etc.
3. With respect to neologism and lexicography, see, for example, discussion by Svensén (1993: 40ff), Landau (2001: 202ff), on headword selection; Hanks (2005: 248-9) on Samuel Johnson, lexical creativity, and institutionalization; and Prinsloo and Gouws (2006) on neologism and longevity.
4. The Bank of English corpus was created by COBUILD at the University of Birmingham. It comprises 450 million words (approximately 71% British English, 21% North American English, 8% Australian English), mainly taken from 1990s written sources but including over 20 million words of spoken interaction and 40 million words of transcribed radio broadcasts.
5. *Veritable* is discussed by Sinclair (2003: 91-103), who draws attention to its metaphorical collocates. Krishnamurthy (1987: 81-85) uses *veritable* to demonstrate compilation procedures on the COBUILD project.
6. The prefix *sub-* can also have this meaning: *sub-Orwellian, sub-Wagnerian* etc.
7. *Mcglitterati* are Scottish glitterati; *lagerati* are supporters of a football team; *twitterati* parallels *chatterati*, but with implications of the worthlessness of their comments.
8. See Kemmer (2003: 73-4, 82ff) for discussion of words in *-erati* from morphological and cognitive perspectives.
9. OALD7 repeats it in its appendix of sayings and proverbs, as it does *out of the frying pan into the fire*.
10. *Silver lining* is a compound headword in *Collins Cobuild English Dictionary* (1995: 2nd edition), which was based on BoE; also in its first edition of 1987.
11. ODEI is a republication, with minor revisions, of *Oxford Dictionary of Current Idiomatic English. Volume 2: Phrase, Clause and Sentence Idioms* (1983).
12. Influenced by a non-standard but fashionable pronunciation of /f/ for /θ/.
13. The same metaphor is found in *keen*: literally 'sharp', metaphorically enthusiastic.
14. Recurrent misspellings such as these, based on mishearings or misinterpretations, are sometimes referred to as 'eggcorns' (Mark Liberman's term (from a mishearing of *acorn* as *eggcorn*): for example, *bare-faced lie, bear-faced lie; foolproof, full proof; lip-sync(h), lip-sing; throes of passion, throws of passion*. See the website <http://eggcorns.lascribe.net>.
15. This can be related to current discussions about whether corpus- and frequency-determined selection of lemmas etc. for dictionaries corresponds appropriately to the kind of information sought by users, as demonstrated through log file records of look-ups: see De Schryver et al 2006, Bergenholtz and Johnsen 2007.

References

Dictionaries

- Cowie, A.P., R. Mackin and I. McCaig. 1993. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press. (ODEI)
- Pearsall, J. (Ed.). 1998. *New Oxford Dictionary of English*. Oxford: Oxford University Press.

- Rundell, M. (Ed.).** 2005. *Macmillan Phrasal Verbs Plus*. Oxford: Macmillan.
- Rundell, M. (Ed.).** 2007. *Macmillan English Dictionary for Advanced Learners*. Second edition. Oxford: Macmillan. (MED2)
- Sinclair, J. and S. Bullon (Eds.).** 1995. *Collins Cobuild English Dictionary*. Second edition. London/Glasgow: HarperCollins.
- Sinclair, J. and R. Moon (Eds.).** 1989. *Collins Cobuild Dictionary of Phrasal Verbs*. London/Glasgow: HarperCollins.
- Sinclair, J. and R. Moon (Eds.).** 1995. *Collins Cobuild Dictionary of Idioms*. London/Glasgow: HarperCollins. (CDI)
- Summers, D., S. Bullon and A. Gadsby (Eds.).** 2003. *Longman Dictionary of Contemporary English*. Fourth edition. Harlow: Longman. (LDOCE4)
- Wehmeier, S. (Ed.).** 2005. *Oxford Advanced Learner's Dictionary of Current English*. Seventh edition. Oxford: Oxford University Press. (OALD7)

Other literature

- Aitchison, J.** 2007. *The Word Weavers: Newshounds and Wordsmiths*. Cambridge: Cambridge University Press.
- Apresjan, J.** 1974. Regular Polysemy. *Linguistics* 142: 5-32.
- Bauer, L.** 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bergenholtz, H. and M. Johnsen.** 2007. Log Files Can and Should Be Prepared for a Functionalistic Approach. *Lexikos* 17: 1-20.
- Carter, R.** 1999. Common Language: Corpus, Creativity and Cognition. *Language and Literature* 8(3): 195-216.
- Carter, R.** 2004. *Language and Creativity: The Art of Common Talk*. London: Routledge.
- Chiaro, D.** 1992. *The Language of Jokes: Analysing Verbal Play*. London: Routledge.
- De Schryver, G.-M., D. Joffe, P. Joffe and 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.
- Hanks, P.** 1987. Definitions and Explanations. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 116-136. London: Collins.
- Hanks, P.** 2005. Johnson and Modern Lexicography. *International Journal of Lexicography* 18(2): 243-266.
- Katamba, F.** 1994. *English Words*. London: Routledge.
- Kemmer, S.** 2003. Schemas and Lexical Blends. Cuyckens, H., T. Berg, R. Dirven and K.-U. Panther (Eds.). 2003. *Motivation in Language: Studies in Honor of Günter Radden*: 69-97. Amsterdam: John Benjamins.
- Kövecses, Z. and P. Szabó.** 1996. Idioms: A View from Cognitive Semantics. *Applied Linguistics* 17(3): 326-55.
- Krishnamurthy, M.** 1987. The Process of Compilation. Sinclair, J.M. (Ed.). 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 62-85. London: Collins.
- Lakoff, G. and M. Johnson.** 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.
- Lindstromberg, S.** 1997. *English Prepositions Explained*. Amsterdam: John Benjamins.
- Maybin, J. and J. Swann.** 2006. *The Art of English: Everyday Creativity*. Basingstoke: Palgrave Macmillan/Open University.
- McCarthy, M.** 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Moon, R.** 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Oxford University Press.
- Nash, W.** 1985. *The Language of Humour*. Harlow: Longman.
- Nunberg, G. and A. Zaenen.** 1992. Systematic Polysemy in Lexicology and Lexicography. Tommola, H., K. Varantola, T. Salmi-Tolonen and J. Schopp (Eds.). 1992. *EURALEX '92 Proceedings. Papers Submitted to the 5th EURALEX International Congress, Tampere 1992*: 387-396. Tampere: University of Tampere.
- Ostler, N. and S. Atkins.** 1991. Predictable Meaning Shift: Some Linguistic Properties of Lexical Implication Rules. Pustejovsky, J. and S. Bergler (Eds.). 1991. *Lexical Semantics and Knowledge Representation*: 76-87. Berkeley: University of California.
- Prinsloo, D.J. and R. Gouws.** 2006. Fashion Words in Afrikaans Dictionaries: A Long Walk to Lexicographic Freedom or Just a Lexical Fly-by-Night? Corino, E., C. Mareello and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th-9th, 2006*: 301-312. Alessandria: Edizioni dell'Orso.
- Sinclair, J.** 2003. *Reading Concordances: An Introduction*. Harlow: Pearson Longman.
- Svensén, B.** 1993. *Practical Lexicography*. Oxford: Oxford University Press.

Traduire les énoncés religieux indigénisés: cas de la traduction du cilubà en français et en anglais*

Germain Mulamba Nshindi, *Institut Supérieur Pédagogique, Mbujimayi, République Démocratique du Congo (ngmulamba@hotmail.com / sgacispmj@yahoofr)*

Résumé: Cet article s'efforce à examiner les difficultés inhérentes à la traduction du cilubà en français et en anglais de certains énoncés, à savoir, les expressions religieuses, les devises et les proverbes. Il a essayé de démontrer qu'une fois que le message biblique a été indigénisé en cilubà, il devient très difficile de le traduire à nouveau dans les langues européennes, en l'occurrence en français et en anglais. En effet, il contient des «africanismes» ou des réalités locales qui sont puisées dans la cosmologie africaine, laquelle n'est pas toujours conforme à la foi chrétienne. Enfin, l'emploi de l'ethnographie de communication du cilubà pendant les cérémonies religieuses incite à l'emploi des métaphores, des proverbes et autres aspects du langage imagé. À moins de recourir à la traduction fonctionnelle, l'interprète ne peut pas aider les locuteurs français et anglais à participer à ces pratiques langagières.

Mots clés: ÉNONCÉ, DEVISE, CONTEXTE, INDIGÉNISATION, RENOUVEAU LITURGIQUE, TRADUCTION, TRANDUCTOLOGIE, PARÉMIES, ETHNOGRAPHIE DE COMMUNICATION, LANGAGE IMAGÉ

Abstract: **Translating Indigenized Religious Utterances: A Case of the Translation of Cilubà into French and English.** This article is an attempt at examining the difficulties inherent in translating certain utterances, namely religious expressions, praise names and proverbs, from Cilubà into French and English. It tries to show that once the Biblical message has been indigenized in Cilubà, it becomes very difficult to translate it back again into the European languages, in this case French and English. In fact, it contains "Africanisms" or local realities which draw on African cosmology which is not always in conformity with the Christian faith. Finally, the use of the ethnography of communication of Cilubà during religious ceremonies leads people to turn to metaphors, proverbs and other aspects of figurative language. Unless turning to functional translation, the interpreter cannot assist French and English speakers to participate in these speech events.

Keywords: UTTERANCE, PRAISE NAME, CONTEXT, INDIGENIZATION, LITURGICAL REVIVAL, TRANSLATION, TRANSLATION THEORY, PROVERBS, ETHNOGRAPHY OF COMMUNICATION, FIGURATIVE LANGUAGE

* Cet article est une version adaptée de notre communication «Le traducteur face au renouveau liturgique» présentée au Colloque Théologique organisé par la Faculté de Théologie Protestante de Mbujimayi, République Démocratique du Congo, 3-5 janvier 2000.

1. Introduction

Cet article essaie d'examiner les difficultés inhérentes à la traduction du cilubà en français et en anglais des énoncés religieux indigénisés couramment utilisés dans les cantiques, les discours et les textes de la liturgie catholique romaine. Le cilubà est une langue bantou parlée dans les deux provinces du Kasai en République Démocratique du Congo. Il est connu sous le sigle linguistique L31, selon la classification de Guthrie (1948).

En effet, parmi les innovations remarquables de ces dernières décennies dans l'Église Catholique Romaine d'Afrique figure «le renouveau liturgique». Il se caractérise entre autres par l'africanisation du rite religieux, l'inculturation du message biblique et linguistiquement, par l'utilisation des langues locales en remplacement du latin et des autres langues européennes. À la suite des sociolinguistes, nous pouvons parler de «l'indigénisation» (Kachru 1982) de la Bonne Nouvelle. Selon Kabasélé (1986a: 270), «inculturer la foi signifie insérer le message chrétien dans une culture, y adhérer avec ses modes de penser, d'agir, de vivre; avec ce qu'on est et aspire à être ...»

Au cœur de cette entreprise se trouvent deux éminences africaines, à savoir le Cardinal Albert Joseph Malula de la République Démocratique du Congo et le Révérend Père Englebert Mveng du Cameroun.

Bien plus, sur le plan de la praxis langagière, ces langues sont utilisées conformément à l'ethnographie de communication spécifique à chaque langue et à chaque culture africaines. En d'autres termes, cette pratique langagière va au-delà de la simple traduction du message biblique européen et devient ancrée dans la créativité africaine pour atteindre le fond de la pensée et de la croyance des africains.

À ce moment où l'interprétation directe ou traduction simultanée est devenue une pratique courante dans la plupart des confessions religieuses, nous avons jugé nécessaire de cogiter sur la traduction du cilubà en français et en anglais de certains énoncés utilisés lors du rite religieux catholique. Notre objectif est d'apprécier le degré d'inculturation de la Bonne Nouvelle dans cette culture africaine et les difficultés potentielles de sa réexpression de ce message indigénisé en français et en anglais.

En effet, sur le plan lexical et discursif, le continuum du registre religieux catholique lubà englobe des mots d'emprunt d'origine latine, des mots altérés ou assimilés en cilubà, et des lexèmes purement natifs n'ayant aucun lien apparent avec les textes européens. Kabasélé fait le même constat dans sa classification des «devises» concernant Jésus, c'est-à-dire, «des syntagmes épithétiques élogieux, plus ou moins figés, qui ... accompagnent ou remplacent le nom du héros» (Faïk-Nzuji 1974: 14). Ainsi identifie-t-il «les titres qui sont puisés dans (ou inspirés de) la Révélation ..., ceux empruntés à la Tradition Bantu et enfin ceux issus soit d'une synthèse entre la Tradition Bantu et la Bible, soit d'une simple coïncidence ou juxtaposition des deux» (Kabasélé 1986a: 264) comme illustré ci-dessous:

- (a) Devises tirés de la Révélation:
Mulami mulenga: Le Bon Pasteur (The Good Shepherd)
Ntambwa wa cisàmbà cyà Yuudà: Le Lion de la tribu de Judée (The Lion of the tribe of Judaea)
- (b) Devises tirés de la Tradition bantu:
Nsànga mweyèmèna Mulopò: Le chêne appuyé sur Dieu (The oak supported by God)
Ngabu wa ncibula bilòbò: Le bouclier sur qui se brisent les héros (The shield on which the brave warriors are broken)
- (c) Devises tirés de la Tradition bantu et la Bible:
Mufwà mwalùka wa Maweja: La mort revenu de Dieu (The Resurrection of God)
Mfùmù, Mukalenga, Mulopò: Chef, Seigneur (Chief, Lord)

Ce continuum révèle aussi les divers degrés de la profondeur d'inculturation des lexèmes religieux en cilubà. Les lexèmes natifs font partie du domaine du langage imagé, langage difficile à traduire en langues européennes à cause de différences culturelles très prononcées. Notre but est de démontrer qu'une fois inculturé, le texte liturgique en cilubà devient difficile à restituer en langues européennes, en français et en anglais pour le cas d'espèce. Le traducteur a besoin d'une bonne connaissance de la cosmogonie lubà et d'une bonne maîtrise des stratégies de traduction. Certes, une telle entreprise nécessite non plus la traduction presque littérale typique à la Bible, c'est-à-dire «la traduction sémantique», mais souvent la «traduction fonctionnelle» ou traduction culturelle (Newmark 1982).

Nous arguons qu'en acceptant l'utilisation des langues africaines dans les rites religieux, on a aussi libéré le langage du croyant, et partant, le croyant lui-même. D'où l'emploi du langage imagé et de l'ethnographie de communication propres à la culture des croyants.

À cette fin, nous esquisserons d'abord la théorie de la traduction et l'ethnographie de communication. Ensuite, nous procéderons à la traduction en français et en anglais des quelques énoncés lubà. Il s'agit principalement des expressions, des devises et des métaphores. Par ailleurs, nous traduirons aussi en français et en anglais quelques expressions lubà relatives à l'ethnographie de communication pour illustrer l'indigénisation du rite religieux par la pratique langagière. La conclusion résumera l'essentiel de notre argument.

2. Considérations méthodologiques

Cette étude focalise à la fois sur la théorie de traduction et sur l'ethnographie de communication. Elle est basée sur des données écrites et orales, à savoir les textes liturgiques, les prières, les cantiques, les prêches et les commentaires des modérateurs des messes. Ainsi, tenterons-nous de traduire un échantillon d'é-

noncés courants dans ces données. Cependant, contrairement aux études traductologiques qui examinent un texte en langue de départ et sa traduction en langue d'arrivée à la recherche de stratégies utilisées dans ce processus, la présente étude table sur les énoncés de la langue d'arrivée pour les retracer dans la langue de départ afin de détecter les déviations intervenues dans cette dernière. En effet, les énoncés du cilubà sont plus élaborés que leurs sources latines et françaises comme conséquence de leur inculturation dans ces dernière langues.

Concernant l'ethnographie de communication, nous nous sommes intéressé aux stratégies langagières utilisées par les officiants, les modérateurs des messes et les chorales pour transmettre d'une façon adéquate le message liturgique d'origine juive aux croyants lubà. Notre approche est purement descriptive en ce qu'elle consiste à expliciter ces messages imagés en rapport avec les textes latins et français d'où ils tirent leur origine pour prouver leur indigénisation.

3. Quelques considérations sur la traduction

La traduction est une vieille discipline dont la longue tradition remonte aux anciennes civilisations, notamment à la culture grecque, si pas plus loin encore dans le temps. Plusieurs définitions de la traduction sont en vogue. Par exemple, pour Taber et Nida (1971: 11), «la traduction consiste à reproduire dans la langue réceptrice le message de la langue source au moyen de l'équivalent le plus proche et le plus naturel, d'abord en ce qui concerne le sens et ensuite en ce qui concerne le style». Bien plus, pour théoriser explicitement sur l'objectif de communication de traduction en fonction de récepteurs précis Nida (1964) a défini deux concepts d'équivalence entre le texte de départ et le texte d'arrivée: l'équivalence formelle et l'équivalence dynamique. La première cherche à reproduire la forme du texte de départ tandis que la dernière cherche à répondre aux besoins du destinataire.

Au niveau théorique, deux paradigmes semblent se dégager dans les différentes écoles de la traduction. Nonobstant les différences terminologiques, les paradigmes se ramènent à la dichotomie entre «la traduction sémantique» et «la traduction fonctionnelle». En fait, selon Kambaja (2007: 50) «la traduction est dite sémantique lorsque elle tente, selon les contraintes syntaxiques et sémantiques de la langue d'arrivée, de reproduire la signification contextuelle précise du texte de départ». Par contre, la traduction fonctionnelle est celle qui vise les équivalences dans l'accomplissement de la même fonction au-delà de différences de formes. Dans le même ordre d'idées, Gonzalez (2003) déclare:

L'équivalence fonctionnelle consiste à atteindre, au moyen des éléments linguistiques, extralinguistiques (ou contextuels) et culturels, le même but, la même fonction que le texte source. Il faudrait donc que le texte cible produise les mêmes actes pragmatiques que le texte source.

Il identifie les types d'équivalences suivantes: l'équivalence linguistique, l'équivalence pragmatique, l'équivalence stylistique, l'équivalence sémantique, l'équivalence formelle, l'équivalence référentielle, l'équivalence paradigmatique, l'équivalence dynamique et l'équivalence fonctionnelle.

À côté de cette terminologie technique existe celle courante qui classe la traduction en «traduction littérale» et «traduction littéraire ou culturelle» qui correspondent respectivement à la «traduction sémantique» et «la traduction fonctionnelle» susmentionnées. En définitive, le débat tourne autour de la fidélité à observer par rapport au texte de départ ou de la langue source.

Il sied cependant de noter que la traduction est avant tout traduction du discours et non traduction des mots pour que l'on puisse s'en tenir aux équivalences sémantiques. La notion d'équivalence implique la comparaison et l'égalité des entités en présence. Les récentes tendances de la traductologie dont la théorie interprétative (Seleskovitch et Lederer 2001) et la théorie de pertinence (Sperber et Wilson 1986) illustrent bien ce débat sur l'objet réel de la traduction. Elles reconnaissent que l'on traduit non pas les langues en présence, mais le sens que l'on a compris comme celui que le contexte confère à l'énoncé. Luscher (2001) parle de la traductologie du discours qui considère l'énoncé comme unité de traduction.

Comme nous le verrons dans l'interprétation des données, l'inculturation du message liturgique amène le traducteur et l'interprète à restituer le discours et non les simples structures grammaticales ou lexicales. Nous pouvons conclure ce débat avec Kambaja (2007: 53) qui stipule ceci: «En tant qu'acte de communication, la traduction ne peut être étudiée que sous un angle pragmatique, prenant en compte les aspects sémantiques, cognitifs, logiques et encyclopédiques.» Salah et al. (2003) résumant mieux ces préoccupations par le titre de leur livre *Traduire la langue, traduire la culture*. Tous ces facteurs sont sous-entendus dans «le contexte» qui définit ainsi la culture.

Sur le plan pratique, la traduction nécessite des stratégies variées pour rendre le message de la langue de départ à la langue d'arrivée le plus fidèlement possible en dépit de l'aphorisme italien «traduttore, traditore», c'est-à-dire, traduire c'est trahir. Selon Eno Belinga (1978: 58), cet aphorisme «établit fatalement que toute traduction est infidèle puisqu'elle trahit la pensée du texte original». Parmi les stratégies le plus en vue, nous pouvons mentionner le chassé-croisé, l'explication, le remplacement métaphorique, l'équivalence dynamique, l'ajout métaphorique, la traduction littérale etc.

La traduction exige du traducteur un effort de créativité dans le choix des termes propres et expressifs et des structures syntaxiques. La qualité de la traduction dépend ainsi du génie créateur du traducteur à manier la langue d'arrivée tant dans sa forme orale qu'écrite.

Comme soubassement théorique, dans les lignes qui suivent, nous explorons certaines sources des difficultés réelles inhérentes à la traduction du langage imagé. Nous essayons d'illustrer comment la même réalité est rendue dans ces trois langues: le français, l'anglais et le cilubà.

Certains énoncés sont difficiles à traduire parce que les langues en présence perçoivent les réalités différemment. Le traducteur doit procéder à la traduction fonctionnelle pour rendre le message compréhensible dans la langue cible même si cela doit entraîner des contradictions apparentes. En linguistique contrastive, tel est souvent le cas en lexicologie pour les termes considérés comme ayant des «sens étranges». Par exemple, l'heure n'est pas comptée de la même façon en kiswahili et dans certaines langues européennes. Ainsi, *9 heures du matin* est «nine o'clock» en anglais mais «saa tatu» en kiswahili, alors que *tatu* signifie «trois» et non «neuf». La raison reste simple: les occidentaux comptent les heures du jour après minuit. Par contre, les waswahili commencent leur journée au lever du soleil. Ainsi se retrouvent-ils en retard de six heures par rapport aux occidentaux.

D'autres exemples peuvent aussi illustrer ce phénomène:

(1)(a) Filer à l'anglaise

se traduit en anglais comme

(b) To take French leave

mais en cilubà, respectivement comme

(c) Kuongoloka (quitter en catimini)

et

(d) Kuongoloka bu kabundi (quitter en catimini comme une genette)

French n'est pas la traduction anglaise de l'adjectif *anglaise*. Il est curieux de constater que les Français et les Anglais s'accusent mutuellement sur la façon de quitter un lieu en catimini. Pour exprimer la même réalité, le peuple lubà recourt à deux énoncés dont l'un est une comparaison avec la genette, un animal réputé rusé dans les fables pour son intelligence.

Dans les proverbes et autres formes de langage figé, là où telle langue utilise comme symbole un animal, telle autre peut utiliser un oiseau, un animal différent, etc.

(2)(a) Il ne faut pas vendre la peau de l'ours avant de l'avoir abattu/tué

(b) Don't count your chickens before they are hatched

(c) Nkàsù bàtu bàyimwèna mu milabì (Les houes, on les voit dans les manches)

La certitude, ou mieux, la méfiance sur les résultats futurs est rendue en français au moyen de l'animal «ours» (bear), en anglais par «chickens» (poules) et en cilubà par «nkàsù» (houes). Pour réussir dans cette entreprise, le traducteur doit recourir à l'équivalence fonctionnelle pour ne pas trahir le message de la langue source.

Bien plus, au niveau de structures des langues, on peut passer de telle partie du discours à telle autre, avec souvent l'élaboration ou la condensation du message:

- (3)(a) Mieux vaut prévenir que guérir
- (b) Prevention is better than cure
- (c) Budimù mbupite bwanga (La prudence dépasse le fétiche)

Le français utilise les verbes (*prévenir* et *guérir*) là où l'anglais utilise les noms (*prevention* et *cure*). Le cilubà recourt aussi aux noms mais donne l'équivalent et non la traduction proprement dite de l'aphorisme français. Le français et l'anglais opèrent au sein du même champs sémantique (*prévenir*, *prevention*) tandis que le cilubà utilise les termes d'un autre domaine (*budimù* et *bwanga*).

Certaines traductions sont basées sur la recherche des mêmes effets dans les deux langues:

- (4)(a) Monsieur Dupond et Monsieur Dupont
- (b) Mr Thompson and Mr Thomson

Il s'agit de personnages des bandes dessinées *Tintin*. Pour produire l'ambiguïté créée par les homophones, le traducteur de l'énoncé français en anglais a fait recours aux noms anglais qui sont aussi épelés différemment mais sont prononcés de la même façon.

Enfin, la traduction des énoncés de langues africaines en langues européennes exige souvent le recours à «la composition» comme procédé linguistique. Celle-ci permet au traducteur de combiner les éléments de deux langues pour exprimer avec précision une idée qui resterait vague si le terme générique était employé. Ainsi, ces devises sont traduites comme suit:

- (5)(a) Dijinda ntunga mùlòngò
- (b) La fourmi-dijinda (la tête de file des fourmis)
- (c) The ant-dijinda (the leader of the line of ants)

Il y a plusieurs espèces de fourmis en cilubà, chacune ayant ses propres habitudes et son propre comportement. Traduire *dijinda* simplement comme *fourmi* créerait de la confusion, ou pire, manquerait de mettre en lumière la qualité de «meneur» ou de «guide» spécifique à cette espèce, qualité que l'on aimerait attribuer métaphoriquement au référent (Jésus Christ pour le cas d'espèce).

- (6)(a) Cinkùnkù nsanga bilembi
- (b) L'arbre-cinkùnkù (rassembleur des chasseurs)
- (c) Cinkùnkù-tree (gatherer of the hunters)

Comme pour le cas précédent, la composition est très importante pour maintenir le terme *cinkùnkù* dans le lexème. Cela permettrait de conserver le sens profond de cette devise, à savoir, la qualité de rassembleur reconnu au référent.

En guise de conclusion, il revient donc au traducteur et à l'interprète de choisir les stratégies de traduction appropriées pour rendre le message plus fidèlement dans la langue d'arrivée.

4. Ethnographie de communication

Un autre aspect visible de l'indigénisation du rite religieux reste sans doute l'adoption de l'ethnographie de communication des cultures africaines. Pour toucher les croyants dans leur moi-profond, le modérateur, l'officiant et même le compositeur des cantiques essayent de rapprocher le message liturgique de la philosophie et culture de ceux-ci. D'où l'emploi du langage imagé, notamment les aphorismes, c'est-à-dire, les proverbes au sens large.

En effet, les proverbes jouent un grand rôle dans le discours africain. Ils sont les signes visibles de l'expression de la sagesse non seulement du sujet parlant, mais aussi de celle de l'auditoire pour lequel le langage est adapté en conséquence. Selon Schmied (1991: 52):

L'une de fonctions du proverbe est de résumer les points qui peuvent être difficiles à expliquer. Bien plus, les proverbes montrent la sagesse du locuteur ... Ils neutralisent ce qui serait autrement des vérités désagréables ou dangereuses pour l'auditeur.

Dans le même ordre d'idées, Oladele (1981: 26) écrit:

Les proverbes concernent tous les aspects de la vie. Ils sont utilisés pour exprimer avec force les paroles des sages et font partie du patrimoine de vieilles personnes, qui les utilisent pour transmettre des leçons morales précises, avertissements et conseils parce qu'ils font un grand impact sur l'esprit plus que les mots ordinaires.

Lors de cérémonies religieuses, les proverbes sont souvent utilisés par le modérateur pour introduire les lectures saintes, mais presque toujours pour les conclure ou les résumer.

Par ailleurs, sur le plan d'interaction, nous assistons à un dialogue entre le modérateur et le public. Le premier pose une question ou prononce une formule introductive de manière à préparer le public sur le sens des lectures saintes ou du message à venir. Comme conséquence, un même texte liturgique peut être introduit ou conclu au moyen de parémies différentes selon la créativité et l'inculturation du modérateur ou de l'officiant, ou selon l'exégèse que chacun d'eux fait de la lecture elle-même.

Mais que représente tout ceci pour le traducteur, se demanderait-on? En fait, bien que le modérateur et l'officiant utilisent un langage simple et ordinaire pour mieux communiquer avec le public, le souci de sonner sage le pousse souvent à employer les parémies. Celles-ci prennent souvent l'interprète au dépourvu car elles exigent une interprétation rapide du message afin

de procéder à la traduction fonctionnelle, à la traduction libre ou à l'explication de l'acte langagier comme en cas d'onomatopées.

Parfois certains animateurs et modérateurs utilisent des parémies n'ayant aucune relation directe avec le message liturgique. C'est-à-dire, le contexte n'est pas approprié pour faciliter la traduction à cause de la discordance entre l'acte posé et le message utilisé pour l'annoncer. Ce qui complique la tâche de l'interprète. À titre illustratif, pour annoncer le port des étoles par les nouveaux prêtres comme symbole de leur nouvelle identité, le prêtre modérateur du jour utilisera cette parémie:

Modérateur:

- (7)(a) Bulenga bwà cibangà
- (b) La beauté du menton
- (c) The beauty of the chin

Public:

- (8)(a) òmwevu
- (b) c'est la barbe
- (c) is the beard

Non seulement le contexte de cet usage n'est pas approprié, mais la véracité même de cette parémie reste relative car le lien entre la barbe et la beauté masculine n'est pas une valeur universelle.

En résumé, grâce au renouveau liturgique, le modérateur, l'officiant et le public ne sont plus relégués au simple rôle des répétiteurs des paroles figées latines dont ils ne comprenaient pas toujours le sens comme dans le cas des formules incantatoires ou magiques. Ils sont devenus créateurs des discours qui utilisent le message liturgique comme simple point de départ et puisent dans la cosmologie et la philosophie lubà dans leur conception des actes langagiers.

5. Essai de traduction des énoncés religieux

Dans cette entreprise, nous fournissons chaque fois un énoncé en cilubà, suivi de sa traduction en français et en anglais. Ensuite, nous ajoutons d'autres énoncés qui attestent l'indigénisation du message et présentent de sérieuses difficultés de traduction pour l'interprète et le traducteur. Certes, l'emphase sera placée sur le message en cilubà pour évaluer comment les traducteurs ont atteint l'africanisation du message liturgique.

- (9)(a) Mwâna wa balùmè umwe
- (b) Le fils unique
- (c) The only Son
- (d) mwâna mulela umwe: un enfant né seul

- (e) mulela umwe: né seul
- (f) Nkônga mikùjà wa Màriyà: rassembleur des fétus de Marie

Pour le cilubà nous enregistrons plusieurs traductions selon les degrés d'inculturation. Comme on peut le remarquer, l'équivalence paradigmatique semble parfaite entre les expressions françaises et anglaises. Nous avons donc deux lexèmes, si nous considérons l'article en anglais et en français comme un élément grammatical sans contenu sémantique.

En cilubà, nous avons l'expression *mwâna wa balùmè umwe*. Sa traduction littérale signifie «un garçon» et non pas «le fils unique». D'où l'emploi de l'expression appropriée *mwâna mulela umwe* «un enfant né seul» trouvée dans le Credo par exemple. Celle-ci est même souvent réduite à la devise *mulela umwe* «né seul» (Kabasélé 1986a: 264) dans laquelle le terme *mwâna* «enfant» est omis ou reste sous-entendu.

Enfin, le cantique utilise la devise *nkônga mikùjà wa Màriyà*, un langage imagé qui témoigne d'une bonne connaissance de la culture lubà par le compositeur. Cette expression signifie littéralement «le fétus qui rassemble tous les fétus de Marie» et connote que cet enfant spécial rassemble en son sein les attributs de tous les enfants potentiels de Marie. Ce qui est une image de richesse et non celle de stérilité ou de pauvreté de fécondité inhérente à la notion de fils unique chez l'africain.

- (10)(a) Kubèènesha
- (b) Bénir
- (c) To bless
- (d) kwela lupèmbà
- (e) kutentekeela byanza

Kubèènesha est un emprunt linguistique français assimilé dans lequel on identifie facilement la racine *ben-* provenant de *bénir*. A part cet emprunt, nous avons aussi *kwela lupèmbà* «asperger avec du caolin», une expression native qui trouve son origine dans les pratiques animistes propres à la plupart de sociétés africaines.

Kubèènesha et *kwela lupèmbà* diffèrent non seulement par leurs origines respectives, mais aussi par la paralinguistique qui les accompagne. En tant que symbole du christianisme, *kubèènesha* est associé avec le signe de la croix alors que *kwela lupèmbà* sous-entend l'aspersion au moyen du caolin ou le dessin au caolin sur le front et les joues du sujet.

Enfin, dans les confessions religieuses où le signe de croix n'est pas de mise, on pratique l'imposition des mains (*kutentekeela byanza*). Néanmoins, même au sein de l'Église Catholique Romaine, cette pratique est courante dans le Mouvement Charismatique et autres dont le rite religieux se rapproche de plus en plus de ceux des Églises de Réveil.

- (11)(a) Kutooka zeezeze

- (b) Être pur
- (c) To be pure/cleansed

Des termes natifs et des emprunts linguistiques existent côte à côte pour des raisons souvent peu évidentes. De plus, la dualité des termes est entretenue parfois même là où l'emprunt linguistique pouvait être évité comme dans ces extraits des cantiques:

- (d) ... ñshaalà zeezeeze_Mfùmù wanyì
... que je sois très pur, Mon Seigneur
... provided that I remain pure, my Lord
- (e) ... ñtookà mupìta nêjè
... que je devienne plus blanc que la neige
... provided that I become whiter than snow
- (f) ... mutòòke bu bùùsù bwà ndânda
... aussi blanc (pur) qu'un flocon de coton
... as white (pure) as a flock of cotton

Dans ces extraits, la notion de pureté est exprimée en terme de «blancheur». Dans le premier cantique (11)(d), nous avons un terme natif (*kuooka*) *zeezeeze* (très blanc). Dans le deuxième (11)(e), nous avons l'emprunt linguistique français *nêjè* (neige). Enfin, dans le troisième (11)(f), nous avons l'image locale *bùùsù bwà ndânda* (flocon de coton). Sa traduction littérale est «flocon de coton», tandis que sa traduction fonctionnelle doit être «flocon de neige», la mesure de blancheur dans la langue française étant attestée par l'expression *blanc comme neige*.

Les parémies présentent plus de difficultés de traduction que les lexèmes simples. Ainsi, là où le dialogue est engagé, l'interprète doit souvent attendre la fin de l'échange dans son contexte afin de concevoir sa traduction:

Modérateur:

- (12)(a) Bààya
- (b) S'ils partent
- (c) If they go

Public:

- (13)(a) Wâya
- (b) Pars (aussi)
- (c) Go (also)

Modérateur:

- (14)(a) Bààshààla
- (b) S'ils restent
- (c) If they stay

Public:

- (15)(a) Washààdilàmu
- (b) Reste pour toujours
- (c) Stay for ever

La traduction littérale ci-dessus n'a pas de sens. En fait, il s'agit ici d'une simple interaction de nature à attirer l'attention du public sur le message important à venir. À la fin de cet échange, l'interprète peut alors exploiter le contexte et procéder à la traduction fonctionnelle au moyen de parémies suivantes pour permettre aux étrangers de saisir le message:

- (d) Il faut hurler avec les loups
- (e) When in Rome, do as the Romans do.

Toutes ces traductions invitent l'assistance à saisir les couleurs locales, c'est-à-dire, à s'adapter au contexte ou à suivre l'exemple des autres membres de la communauté que l'on rejoint.

D'autres parémies, bien que traduisibles, nécessitent des explications supplémentaires pour rendre le message explicite:

Modérateur:

- (16)(a) Cimanyinu cyà bânà bàà Kabòngù
- (b) Le signe distinctif des enfants de Kabòngù
- (c) The distinctive sign of Mr Kabòngù's children

Public:

- (17)(a) ñkabùkì mu nshìngù
- (b) c'est le pendentif au cou
- (c) is an amulet on the neck

Cette parémie a été utilisée à diverses occasions pour annoncer le port des étoles par les nouveaux prêtres et diacres, des chapelets par les nouveaux baptisés et premiers communiantes et des croisettes par les jeunes de lumière promus. Les modérateurs voulaient chaque fois attirer l'attention du public sur l'acte que l'officiant allait exécuter bientôt pour conférer une nouvelle identité aux lauréats.

Pour faire participer les étrangers à ces actes langagiers, l'interprète peut tout au plus leur expliquer la cérémonie en ces termes:

- (18)(a) Maintenant l'officiant va procéder à ...
- (b) Now the priest is going to ...

Certaines parémies, quoique traduisibles littéralement, requièrent une traduction fonctionnelle pour révéler la sagesse dans la langue cible:

Modérateur:

- (19)(a) Nzùbù mukòsòlòka ...
- (b) Une maison divisée ...
- (c) A house divided against itself ...

Public:

- (20)(a) Cyèndà ñkabùtù
- (b) La chute est son sort
- (c) To fall is its fate

Ces traductions peuvent être améliorées au moyen de parémies françaises et anglaises comme suit:

- (d) L'union fait la force
- (e) United we stand, divided we fall.

Une certaine catégorie de parémies ne se prête à aucune traduction. Il s'agit particulièrement des onomatopées. L'interprète peut tout au plus en expliquer le sens ou l'usage:

Modérateur:

- (21) Petètètèpètè èè ...
- (Onomatopée)
- (Onomatopoeia)

Public:

- (22)(a) Patwàkashùla dîyî
- (b) Là où nous avons laissé la voix
- (c) There where we left the voice

La traduction fonctionnelle en français et anglais est respectivement:

- (d) Notre convention était ...
- (e) Our agreement was ...

Le modérateur utilise une onomatopée tandis que le public répond littéralement «là où nous avons laissé notre convention», c'est-à-dire, «notre convention était ...». L'interprète peut expliquer à ses hôtes cet acte en ces termes:

- (23)(a) Il leur rappelle la convention concernant ...
- (b) He/She reminds them of the agreement about ...

Dans leur traitement de la notion d'équivalence, Vinay et Darbelnet (1958: 52) ont abordé le cas des onomatopées en spécifiant que «les deux textes rendent compte d'une même situation en mettant en œuvre des moyens stylistiques et structuraux entièrement différents». Ainsi illustrent-ils le cri de douleur par

Aie! en français et *Ouch!* en anglais, auxquels nous pouvons ajouter *Eyi/Ayi!* en cilubà. *Petètèpètè èè* est une formule introductive d'une interaction qui n'a pas d'équivalent dans l'ethnographie de communication du français et de l'anglais.

Enfin, la cosmologie lubà nous mène parfois dangereusement à la limite entre le christianisme et le paganisme; ce qui rend la traduction bizarre. Par exemple, traduire la devise

(24)(a) Mwanza-nkòngòlò lukànda mvùla wa mudimbì

comme

- (b) L'arc-en-ciel qui arrête la pluie torrentielle
- (c) The rainbow which stops the torrential rain

serait ériger un simple élément cosmique en une divinité ou force cosmique guerrière. Et pourtant la Bible reconnaît l'arc-en-ciel tout simplement comme le signe de l'alliance de paix entre le Seigneur et Noé après le Déluge (Genèse 9: 13). Ainsi, les devises de Jésus, de Dieu ou de la divinité en général sont tirées de divers domaines et associées avec toutes les forces de la nature dont les arbres, les animaux, les oiseaux, les êtres mythiques et mythologiques, les phénomènes cosmiques et atmosphériques etc. (voir Kabasélé 1986a, 1986b).

L'accumulation des devises est fréquente dans les cantiques où les compositeurs louent Jésus Christ en invoquant ses divers attributs ensemble. Cette strophe d'une version de Gloria en est bien un exemple significatif:

(25)(a) Mukalenga wa diulu ...

- (b) Le Seigneur du Ciel
- (c) The Heavenly King

(26)(a) Mwena ngulu yônso ...

- (b) Le Tout-Puissant
- (c) The Almighty

(27)(a) Ntùita lwàbanya màkàlèngà ...

- (b) Ntùita, le donneur du pouvoir
- (c) Ntùita, the giver of power

(28)(a) Ngabu wetù wa kweyemena

- (b) Notre bouclier sur lequel s'appuyer
- (c) Our shield on which to lean

(29)(a) Mwâna mukòkò ...

- (b) L'agneau de Dieu ...
- (c) The Lamb of God ...

(30)(a) Mâyì a kowa ... àdì àtòòkesha bantu

- (b) L'eau à laver ... qui purifie les gens

(c) The water used for washing ... which cleanses the people

(31)(a) Mulopò cyenda ne bantu

(b) Dieu qui chemine avec le peuple (Dieu, le guide du peuple)

(c) God who walks with his people (God, the people's guide)

Toutes ces devises se traduisent simplement par le référent «Jésus Christ», et sont choisies selon le contexte et l'attribut que l'on aimerait mettre en lumière.

En bref, traduire un énoncé liturgique du cilubà en langues européennes requiert moins la connaissance de l'équivalent de chaque mot que celle de la correspondance de la réalité entre les langues et les cultures, de que Taber et Nida (1971) ont qualifié de «traduction pragmatiquement équivalente». En effet, une tâche difficile pour l'interprète du cilubà qui ne dispose pas d'assez de temps pour rendre fidèlement l'énoncé dans la langue cible.

6. Conclusion

Notre objectif dans cette étude consistait à mettre en évidence les difficultés de traduire les énoncés religieux du cilubà en français et en anglais. Ensuite, notre recherche a également voulu démontrer qu'en libéralisant l'utilisation des langues locales dans le rite religieux, le renouveau liturgique catholique a aussi libéré le croyant lui-même dans le domaine langagier. Plutôt que de rester lié à la structure et au contenu du message liturgique tel que le traduisent les écritures saintes, le croyant mulubà en l'occurrence, utilise, par contre, ces messages comme source d'inspiration pour exprimer sa foi selon sa propre culture. Tel est surtout le cas pour les cantiques et les discours des officiants et des modérateurs des messes qui sont parsemés de devises et de langage imagés. Le croyant intègre à la Bonne Nouvelle la cosmologie de sa culture et l'ethnographie de communication de sa propre langue. Dans cet effort d'indigéniser le message liturgique, le traducteur et l'interprète sont obligés d'utiliser la traduction fonctionnelle afin de respecter et de sauvegarder les normes communicationnelles de la langue d'arrivée qu'est le cilubà. Par conséquent, le message devient tellement inculturé qu'il est difficile de le restituer encore fidèlement aux locuteurs des langues européennes à cause des africanismes qu'il contient désormais.

Sur le plan de l'ethnographie de communication, le croyant africain peut bien accueillir le message liturgique apporté par l'européen, sans pour autant renoncer à sa propre culture. Tout comme les juifs déportés à Babylone qui demandaient à leurs détracteurs «Comment chanterions-nous les cantiques de l'Éternel sur une terre étrangère?» (Ps 137: 4), les croyants lubà du renouveau liturgique catholique doivent s'être également demandé, *mutatis mutandis*, «Comment louerions-nous notre Créateur Mawêja-Nààngila, notre Mikombe-a-Kalèèwu dans un langage étranger?» D'où le recours au langage imagé parsemé des devises et des proverbes, et à la paralinguistique propre au cilubà.

Rendre toute cette richesse culturelle en langues européennes susmentionnées est une tâche ardue pour l'interprète et le traducteur.

Bibliographie

- Eno Belinga, Samuel-Martin.** 1978. *Comprendre la littérature orale africaine*. Yaoundé: Éditions St Paul.
- Faïk-Nzuzi, Clementine M.** 1974. *Kàsàlà, chant héroïque lubà*. Lubumbashi: Presses universitaires du Zaïre.
- Gile, Daniel.** 2007. *La traduction. La comprendre, l'apprendre*. Paris: Presses universitaires de France.
- Gonzalez, Gladys.** 2003. *L'équivalence en traduction juridique: Analyse des traductions au sein de l'Accord de libre-échange Nord-Américain (ALENA)*. Thèse de doctorat. Québec: Université Laval, Département des langues, linguistique et traduction. <http://www.theses.ulaval.ca/2003/21362/21362.html> [26 March 2008].
- Guthrie, Malcolm.** 1948. *The Classification of Bantu Languages*. Oxford: Oxford University Press.
- Kabasélé, François.** 1986a. Christ dans l'actualité de nos communautés. Luneau, René, Joseph Doré et François Kabasélé (Éds.). 1986. *Chemins de la christologie africaine*: 263-272. Paris: Desclée.
- Kabasélé, François.** 1986b. Le Christ comme chef. Luneau, René, Joseph Doré et François Kabasélé (Éds.). 1986. *Chemins de la christologie africaine*: 109-125. Paris: Desclée.
- Kachru, Braj B.** 1982. *The Other Tongue: English across Cultures*. Urbana, Ill.: University of Illinois Press.
- Kambaja, M.** 2007. *Approche pragmatique de la traduction français-cilubà. Contribution à la théorie traductologique*. Mémoire de D.E.S. Lubumbashi: Université de Lubumbashi, Département des sciences du langage.
- Luscher, Laure A.** 2001. *Vers une théorie sémanco-pragmatique pour la traduction*. Göppigen: Kümmerle Verlag.
- Newmark, Peter.** 1982. *Approaches to Translation*. Oxford: Pergamon.
- Nida, Eugene A.** 1964. *Toward a Science of Translating with Special Reference to Principles and Procedures Involved in Bible Translating*. Leyde: E. J. Brill.
- Oladele, Taiwo.** 1981. *An Introduction to West-African Literature*. Nairobi: Nelson.
- Salah, Mejri, Bacchouche Taieb et André Class (Éds.).** 2003. *Traduire la langue, traduire la culture: rencontres linguistiques méditerranéennes*. Paris: Maisonneuve et Lerose.
- Schmied, Joseph.** 1991. *English in Africa: An Introduction*. Londres: Longman.
- Seleskovitch, Danica et Marianne Lederer.** 2001. *Interpréter pour traduire*. 4ème édition. Paris: Didier Érudition.
- Sperber, Dan et Deirdre Wilson.** 1986. *La pertinence: communication et cognition*. Paris: Édition de Minuit.
- Taber, Charles et Eugene A. Nida.** 1971. *La traduction: théorie et méthode*. Londres: Alliance Biblique Universelle.
- Vinay, Jean-Paul et Jean Louis Darbelnet.** 1958. *Stylistique comparée du français et de l'anglais: méthode de traduction*. Paris: Didier.

The Effect of Lexicographical Information Costs on Dictionary Making and Use*

Sandro Nielsen, *Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (sn@asb.dk)*

Abstract: An important aspect for any lexicographer to consider is the ease with which users will be able to acquire the necessary information from the data presented in the dictionary, whether electronic or printed. Furthermore, this aspect has high priority with users and reviewers. A distinction is proposed between two general types of lexicographical information costs. Firstly, search-related costs are the effort required by the look-up activities users have to perform when consulting a dictionary to find access to the data they are searching for. It is argued that the access route, article structure, data distribution and cross-references may influence search-related information costs positively as well as negatively. Secondly, comprehension-related costs are the effort connected to the user's ability to understand and interpret the data presented in a dictionary. In other words: How easy or difficult is it for users to understand the data presented? Examples show how textual condensation, dictionary functions and use-situations may impact on the level of comprehension-related information costs. It is thus possible to establish a framework for evaluating lexicographical information costs for the purpose of dictionary making, use, analysis and research.

Keywords: LEXICOGRAPHY, DICTIONARY, INFORMATION COSTS, ACCESS ROUTES, DISTRIBUTION STRUCTURE, CROSS-REFERENCES, TEXTUAL CONDENSATION, DICTIONARY ARTICLES, SEARCH ZONES, DATA, INFORMATION, DICTIONARY FUNCTIONS, USE-SITUATIONS

Opsomming: Die uitwerking van leksikografiese inligtingskoste op die maak en gebruik van woordeboeke. 'n Belangrike aspek vir enige leksikograaf om in ag te neem is die gemak waarmee gebruikers in staat sal wees om die nodige inligting van die data in die woordeboek, hetsy elektronies of gedruk, te verkry. Verder geniet hierdie aspek hoë voorkeur by gebruikers en resensente. 'n Onderskeid word voorgestel tussen twee algemene soorte leksikografiese inligtingskoste. Eerstens, soekverwante koste is die inspanning wat gepaardgaan met die opsoekaktiwiteit wat gebruikers moet verrig by die raadpleging van 'n woordeboek om toegang te verkry tot die data waarna hulle soek. Daar word geredeneer dat die toegangsroete, artikelstruktuur, dataverspreiding en kruisverwysings soekverwante inligtingskoste sowel positief as negatief kan beïnvloed. Tweedens, begripsverwante koste is die inspanning verbonde aan die gebruiker se vermoë om die gegewe data in 'n woordeboek te verstaan en te interpreteer. Met ander woorde: Hoe maklik of moeilik is dit vir gebruikers om die gegewe data te verstaan? Voor-

* This article is partially based on a presentation made at the Thirteenth International Symposium on Lexicography at the University of Copenhagen, Copenhagen, Denmark, 3–5 May 2007.

beelde toon hoe tekstuele verdigting, woordeboekfunksies en gebruikssituasies 'n uitwerking kan hê op die vlak van begripsverwante inligtingskoste. Dit is derhalwe moontlik om 'n raamwerk te skep vir die beoordeling van leksikografiese inligtingskoste vir die doel van die maak, gebruik, ontleiding en ondersoek van woordeboeke.

Sleutelwoorde: LEKSIKOGRAFIE, WOORDEBOEK, INLIGTINGSKOSTE, TOEGANGSROETES, VERSPREIDINGSTRUKTUUR, KRUISVERWYSINGS, TEKSVERDIGTING, WOORDEBOEKARTIKELS, SOEKSONES, DATA, INLIGTING, WOORDEBOEKFUNKSIES, GEBRUIKSITUASIES

1. Introduction

When we consult dictionaries we are in effect seeking knowledge, but acquiring knowledge involves costs, whether we are aware of it or not. The costs involved in searching for data and in reading texts will always be present but are often so insignificant that they go unnoticed. Authors may have adapted their writings to a particular audience and to specific types of reading situations. It is not uncommon to find children's editions of books originally written for adult readers, which have been adapted for children in a language they can understand, for instance the *Encyclopædia Britannica 2007 Children's Encyclopedia*. The publishers have realized that children up to a certain age will not be able to read and understand the adult edition of the encyclopedia properly; children do not consult the *Encyclopædia Britannica* for adults because it is too difficult to read and understand.

We have all been in a situation where, after having consulted a dictionary, we feel let down because the dictionary did not provide the expected help. One of the reasons for our unhappiness with the result of our consultation may be that we did not acquire the information we hoped we would gain by looking up a word in the dictionary. The lack of knowledge development is usually caused by a range of factors, and it is important to be aware of these when working with dictionaries. In this article, I will look at some of the factors that may frustrate the process of consulting dictionaries. This involves an examination of the communicative role of dictionaries and the laying of a theoretical foundation of the relation between their data and the information users may acquire. I will then address factors that may lead lexicographers to choose more carefully between several options when making relevant decisions for the planning and compilation of a dictionary.

2. Dictionaries in the communication process

Dictionaries may be regarded as elements in a communicative framework. Lexicographers have written a number of texts collected in individual dictionaries for the purpose of helping users by sharing knowledge. Bergeholtz, Tarp and Wiegand (1999: 1763) suggest that a dictionary is not a text in itself but

rather a collection of text types or text genres with different communicative functions. It can be argued that the communication process starts when users formulate questions to which they seek answers in their dictionaries and ends when they have found the answers or not. However, it is important to note that users often see the texts in a dictionary as typical examples of a one-way communication: lexicographers communicate data to dictionary users but do not engage in an instant dialogue, nor intend to do so. This is wholly true of printed dictionaries where no one expects lexicographers and users to enter into dialogues during ongoing acts of consultation. Some dictionaries invite users to contact lexicographers or publishers if they have any comments, but this does not form part of specific consultative acts in which users seek immediate help.

It is debatable whether electronic dictionaries allow users to enter into dialogues with lexicographers. Many intranet and internet dictionaries allow users to contact their lexicographers, but whether they receive a reply, and, if so, when, is uncertain. At present, it is fair to say that no electronic dictionaries allow users to enter into direct dialogues with lexicographers in specific instances of consultation so that, during an ongoing act of consultation, users can ask clarifying questions and receive answers immediately. If dialogues are established they will occur after the consultative acts have ended. The future development of interactive online systems may change the dictionary environment, but until then, users are unlikely to regard dictionary consultation as a two-way communication. Lexicographers should take this restricted communicative framework into consideration when planning and compiling dictionaries in order to help users.

3. What are lexicographical information costs?

The costs of acquiring information are generally a natural part of the process and therefore we do not notice them. However, if they do not form part of the natural process they interrupt it, thereby becoming salient, causing users to notice them. In order to establish a workable basis for analyzing the ease or difficulty facing users of dictionaries in their attempts to find help, it is necessary to determine the types of costs involved. Economists were some of the first to use and analyze the concept of information costs, applying it in business economics and in management theory. Example 1 contains a general definition:

information cost

The cost to an organisation of obtaining knowledge of its business environment.

Example 1: Article from *Routledge Dictionary of Economics* (2002).

At first glance, the emphasis in Example 1 appears to be on costs measured in monetary terms, but the concept is somewhat wider. As pointed out by Casson

(2000: 117-122), information costs include what may be called *temporal costs*, i.e. the time spent searching, looking up and understanding data in addition to *pecuniary costs* incurred in appraising and planning activities, exchanging information, and verifying data. Even though this perspective has an economic bias, information costs are also relevant in non-economic settings.

The concept of information costs has come to be used outside economic and management contexts, for instance in information theory, which is reflected in the following definition:

information costs

The costs of acquiring, processing, and using information, part of the costs associated with any type of decision, including a decision how to vote.

Example 2: Article from *Texas Politics* (2007).

Example 2 contains two important points: (a) information costs are part of acquiring information; and (b) information costs are part of processing the relevant data. The definition introduces a third element, which may be referred to as *intellectual costs*, i.e. the mental efforts involved in processing the data. Applied to dictionary consultation, this means that it involves costs to obtain the assistance sought from dictionaries, and to decode the data for finding a useful interpretation.

Lexicographers should carefully consider the ease with which users will be able to find help from their dictionaries. One way of dealing with this issue is to attempt to adapt the concept of information costs to lexicography because dictionaries are carriers of texts. For the purposes of dictionary compiling, use, analysis and research, Nielsen (1999: 111) defines *lexicographical information costs* as the effort that a user believes or feels is associated with consulting a dictionary, an article or any other text part of a dictionary. For the user the important point is the relation between the anticipated information costs and the anticipated information value, i.e. the user's gain from consulting a dictionary in a given consultative act.

Lexicographical information costs may be divided into two distinct types. First, *search-related information costs* are the costs (i.e. effort) related to the look-up acts users have to perform when consulting a dictionary in order to gain access to the data for which they are searching. The more acts the user has to perform to reach the target, the more significant will be the costs, and so will be the risk of going wrong at one or more steps in the search process. The very design and structure of the dictionary may contribute to keeping lexicographical information costs at a low level, as an inappropriate design and structure may lead to high or increasing information costs. De Schryver (2003: 173-182) provides a good account of search options in electronic dictionaries, and typical examples of situations involving search-related information costs comprise: cross-references included without any clear purpose; the number of individual

steps necessary in the look-up process (in printed and electronic dictionaries) before the sought goal is reached; the need for scrolling the screen to find a specific item; and the possibility of searching for a specific word instead of scrolling a long list.

The search options available in electronic dictionaries directly affect the users' perception of costs associated with finding help to solve their problems. It is important to users whether they can only search for full, lemmatized terms, whether they can also search for parts of lemmatized terms (e.g. 'begins with', 'ends with' and 'includes'), or whether they can search for full and partial terms in any part of dictionary articles (e.g. in definitions and in collocations) as well as in the outer texts, such as subject-field components, style guides and different kinds of tables. The above examples may lead some to conclude that all lexicographical information costs are inherently quantitative and temporal in nature, but this is not the case.

The second type of lexicographical information costs may be described as having a qualitative nature. *Comprehension-related information costs* are the costs (i.e. effort) related to the user's ability to understand and interpret the data presented in a dictionary. In other words: How easy or difficult is it for users to process the data in the dictionary? This is directly related to the factual and linguistic competence of users and the way in which the data are presented. The wording and presentation of data in the articles, for instance a high degree of textual condensation in definitions, may increase the information costs. It is clear from the above description of lexicographical information costs that the concept relates to the search for and interpretation of texts or text parts in dictionaries, but it is relevant to determine its sphere of application.

The scope of application is generally the dictionary as a whole, i.e. the entire collection of text types or text genres. However, lexicographical information costs do not directly apply to all of these text components. Tarp (1998: 122-131) distinguishes between *lexicographical components*, which are related to the use and functions of dictionaries, and *extra-lexicographical components*, which are not related to the use and functions of dictionaries. Examples of components irrelevant to the use and functions of dictionaries are acknowledgements, picture credits, eulogistic forewords (such as that found in the *Oxford Advanced Learner's Dictionary*), and single user licence agreements for accompanying CD-ROMs (such as that found in the *Longman Dictionary of Contemporary English*). In the context of lexicographical information costs, only lexicographical components warrant attention, the following being some of the most important:

- Front matter: user guide, subject-field component, writing guide
- Wordlist: articles, inserts
- Back matter: appendices

It follows from this that lexicographers should see their dictionaries as collections of interrelated texts. The above list uses the traditional terminology appli-

cable to printed dictionaries, but this does not exclude electronic dictionaries. On the contrary, authors and users of CD-ROM, intranet and internet dictionaries are also affected by lexicographical information costs as these types of reference works normally each contains several lexicographical components, though their actual positions in relation to the wordlist are difficult to determine. As argued in Nielsen (2006: 10-18), this means that outer texts such as the user guide should be easy to read and actually show how users may find the data presented in the dictionary. For instance, a user guide containing a balanced combination of visual keys and supporting text is usually easier to read and understand than one written in a condensed, straight-text mode. Similarly, the user guide should explain why the different lexicographical components are included, how they supplement each other, and how the components support the functions of the dictionary.

The discussion so far indicates that the use of dictionaries is not cost free and that it is impossible to eliminate information costs. These costs, however, can be minimized. Furthermore, dictionaries are not merely wordlists with adjunct texts but collections of interrelated texts supporting use and functions. Finally, it should be appreciated that a user's choice of dictionary depends on the relation between the expected information costs and the expected informative value when consulting a dictionary. Every time we consult dictionaries for help, we usually have a favourite dictionary, which we consult before any other. The question then is: what is actually in the dictionary?

4. An important distinction between data and information

In our daily lives, we normally use the two words *data* and *information* interchangeably. Nevertheless, when we examine, describe, use and analyze dictionaries in an academic or a professional context it is important to distinguish between the two. As argued by Wiegand (2000a: 22-25; 2005: 216), a dictionary does not contain information, but data that users can convert into information through a mental process and thereby satisfy their needs for help in a given situation. This distinction is also relevant to the discussion of lexicographical information costs, because it sheds light on the way in which users read and understand what is available in dictionaries. Many people have attempted to define data, Example 3 providing a generally accepted definition.

data

any representation such as characters or analog quantities to which meaning is, or might be, assigned.

Example 3: Article excerpt from *Dictionary of Banking and Financial Services* (1985).

This definition indicates that data items do not carry information, but that meaning may be assigned to them. According to Liu (2000: 1), 'information is

processed data that has meaning to its users' and 'all information is "carried" by signs of one kind or another'. By analogy, this means that lexicographical data may be treated as semiotic signs carrying information. When this is assumed, several theoretical approaches are available. It is difficult to say which is the right approach, but Peirce (1992) offers one that specifically deals with the process of assigning meaning to signs through the interplay between what is referred to as universal categories. Peirce (1992: 296) describes the three universal categories called firstness, secondness and thirdness as follows:

First is the conception of being or existing independent of anything else. Second is the conception of being relative to, the conception of reaction with, something else. Third is the conception of mediation, whereby a first and second are brought into relation.

Merrel (2000: 32-33) applies the universal categories to various genres, including poetry and art, on the premise that firstness is possibility, secondness is actuality, and thirdness is potentiality, probability or necessity. Based on his application of Peirce's universal categories to texts, the following is an attempt to describe how meaning is assigned to lexicographical data:

- (a) A sign is a representation on paper/screen and has the possibility of meaning something to someone.
- (b) The actual reading of the sign by the dictionary user creates one or more relations between the sign and the user/reader triggering something in the user's mind.
- (c) The user's reading of the sign and his mindset, his experience and conventions combine to establish a meaning of the lexicographical sign.

This description of data as signs involves two important aspects. Firstly, it can help to understand why different users may interpret the same text or text part differently, depending on their mindsets, experience and the conventions they know and apply. Secondly, data items expressed as signs may take different forms, for instance numerical figures, alphabetical letters, strings of graphemes, graphical representations, symbols, icons and textual suprasegments. I suggest that lexicographers should take both aspects into consideration when writing their dictionary texts with a view to easy consultation. By their very nature, lexicographical information costs cannot be eliminated, but it is appropriate to look at some factors that may help to keep them at a reasonable level.

5. Article structures may reduce information costs

The access route to the data in dictionaries may reduce or increase lexicographical information costs. As shown in section 3, the wordlist and outer texts fall within the scope of information costs and the individual lexicographical

components belong to different text genres. This means that the access structure applying to one component may differ from that applying to another. The central component of a printed dictionary is the wordlist and the end result of a successful search is a text, usually a dictionary article. Most CD-ROM and many online dictionaries are exclusively or substantially based on printed dictionaries whose data have been transferred to electronic media. Consequently, the outcome of a successful search in these dictionaries is usually also a text in the form of an article. Advanced online dictionaries may not have wordlists but databases in which the search is conducted electronically and the result of a successful search may be a screen representation of data, including photographs and video footage, collected from various places in the databases. Nevertheless, it seems reasonable to say that many search results are presented as text in online dictionaries, though perhaps not as dictionary articles proper.

Users place easy and quick access to data items in articles high on their list of priorities. By using standardized article structures in dictionaries, lexicographers make it easy for users to familiarize themselves with the layout of articles. They may divide the text of an article into fields, or zones, that each contains a specific type of data. Users can easily navigate their way through such standardized articles because it involves less effort or incurs a lower cost to find what they are looking for, or to realize that what they are looking for is not in the article.

Each article in a dictionary is different from all other articles, but they may be analyzed using the same approach to access. Wiegand (2000b: 259 et seq.) suggests that articles may be analyzed in terms of search zones, which are relevant when considering search-related information costs. This may be illustrated on the basis of the following article:

final dividend

⟨a, the, -s⟩

endeligt udbytte

⟨et; det endelige udbytte, endelige udbytter, de endelige udbytter⟩

Definition

Final dividend is remaining dividend payable at the end of the financial year. The final dividend constitutes the difference between any interim dividend paid during the year and the total dividend declared at a company's annual general meeting.

Collocations

- declare final dividend of 10%
deklarere endeligt udbytte på 10%
- recommend a final dividend of 9.5%
foreslå et endeligt udbytte på 9,5%
- entitled to the 2006 final dividend
berettiget til endelig udbytte for 2006
- proposed final dividend
forslag til endeligt udbytte

[See also:](#)
interim dividend

Example 4: Article from *English–Danish Accounting Dictionary* (2005-2008).

This article comes from an online English–Danish accounting dictionary, hence the somewhat expanded layout. A closer study shows that it has been divided into separate search zones based on the types of data they contain. The data should be presented as separate search zones within which the user can easily look for individual data items. The way to create search zones in a dictionary article is to make sure that the article text does not appear as a long, unbroken chain of words, signs, etc. This may be achieved by using various structural indicators as signposts such as blank space at the end of a line, typographical symbols, icons, italics and boldface. One result is that users may save time by proceeding directly to the relevant search zone without having to go through the article from the beginning until they reach the zone containing the data they are looking for.

The search zone structure in Example 4 is vertical because all search zones are placed so that they are either above or below one or more other search zones. The first search zone contains the lemma and the item showing the inflectional paradigm; the suprasegmental feature boldface is used as a search zone indicator. Search zone two is made up of the Danish equivalent and the item indicating the inflectional paradigm; boldface is again used as a search zone indicator. The third search zone contains the English definition and is introduced by the headline 'Definition' in a different colour (in this case blue). Search zone four presents the English collocations and phrases and their Danish translations. Again a headline 'Collocations' in a different colour is used as search zone indicator. The fifth search zone has a cross-reference to another lemma, the headline 'See also' in blue functioning as search zone indicator. Users can click on the cross-reference address and be linked directly to the article **interim dividend**. The blank space after the last line of the zone and a blank line between zones mark the end of each search zone.

Some of the search zones in Example 4 are divided into subzones that may facilitate access to the data and reduce temporal costs. For example, the first search zone is divided into two subzones, the first consisting of the boldface lemma followed by the grammatical data within angle brackets. Search zone four contains collocations and phrases in four separate subzones, each introduced by a bullet and with the English collocations against a grey background as signpost. Finally, the overall search path users have to follow consists of the search zone indicators, is positioned vertically and is margined to the left of the article. If the text had been written as a long, unbroken chain of words, signs, etc., the search path would have been sinuous. Even though the article in Example 4 comes from an online dictionary, this description shows that it is possible to analyze the elements of articles in printed and electronic dictionaries in terms of textual topology, search zones and subzones and to arrange

these elements in such a way as to ensure the continuity of the article text and easy access to the data.

Search-related information costs are related to the lexicographical signs introduced in section 4. Signs may appear in various shapes and forms but their meanings depend on the knowledge and experience of readers as well as the conventions they know and apply. For instance, the use of bullets to signpost the beginning of a collocation or phrase is a convention established by the lexicographers of the *English–Danish Accounting Dictionary* and has to be learnt by users of this particular dictionary. Other lexicographers may employ other signposts in their dictionaries. A blank at the end of search zones is a more widely used sign in dictionary articles and may be said to be a generally applicable convention. The use of a suprasegment, for instance boldface, to signpost the lemma is perhaps a universal convention and known by all users of dictionaries. Users generally widen their experience as they consult different dictionaries and thereby familiarize themselves with common and universal signs and conventions, whereas dictionary-specific signs and conventions will have to be learnt on a case-by-case basis. By properly using such signposts in articles, lexicographers can keep the search-related information costs at a low and reasonable level.

6. Dictionary functions and use-situations as benchmarks

Properly arranged textual structures may well facilitate the access to lexicographical data but they do not guarantee the easy understanding of data. The study of comprehension-related information costs generally involves what may be called a qualitative approach. Users reading the data presented in dictionaries need to process these data into information useful to them. The question is what factors to take into account when studying the various reasons causing the information generating process to be more or less costly. One way in which to analyze and evaluate information costs is to relate the data to the functions of dictionaries. In order to do this, it will be necessary to take a closer look at the functions of dictionaries and the needs of users. The function of dictionaries is to provide assistance to particular user groups with specific characteristics in order to cover the needs that arise in specific types of use-situations. The types of use-situations allow lexicographers to distinguish between two overall functions.

Bergenholtz and Nielsen (2006: 286) suggest that dictionaries may be designed to deal with two general types of functions. Firstly, *communication-orientated functions* are those that assist users in solving problems related to the understanding of texts, the production of texts and the translation of texts. For example, in an ongoing written or spoken act of communication between two or more persons a communication-related problem arises and the interlocutors consult a dictionary in order to find help. Secondly, *cognition-orientated functions* are those that provide general and encyclopedic information, special informa-

tion about a topic, or information about language in general or in specific contexts. Unrelated to any ongoing communication situation, a person may want to obtain additional information on a topic such as general cultural information or specialized information concerning an issue within a particular subject-field. In order to find help in such situations, we often consult dictionaries because we think that they can help us solve our problems.

In order to meet the needs of users in specific types of use-situations, lexicographers should consider profiling user groups. Bergenholtz and Nielsen (2006: 285) provide the following list of characteristics lexicographers may take into consideration in an attempt to match user needs and dictionary functions:

- Which language is their native language?
- At what level do they master their native language?
- At what level do they master a foreign language?
- How extensive is their experience in translating between the languages in question?
- What is the level of their general cultural and encyclopedic knowledge?
- At what level do they master the special subject field in question?
- At what level do they master the corresponding LSP in their native language?
- At what level do they master the corresponding LSP in the foreign language?

There may be other relevant types of characteristics for a particular dictionary which do not all apply to all dictionaries, but those listed above are the most important for profiling a specific user group. The important point is that the relation between dictionary functions, user needs and use-situations may serve as benchmarks for assessing comprehension-related information costs, because they can act as standards against which to compare the costs involved.

The real needs of users directly affect the practical work of lexicographers. Laypersons using specialized dictionaries have characteristics, i.e. factual and linguistic competences, differing from those of experts using the same or similar dictionaries. One way of dealing with this is to present the data in ways adapted to the intended user groups. As indicated in section 3, this applies to all lexicographical components. Bergenholtz and Kaufmann (1997: 115-116) illustrate how definitions may be adapted to the user profile of an electronic dictionary of gene technology.

One advantage of electronic dictionaries is that they may allow users to select particular functions for specific groups of users. A layperson having difficulties understanding a newspaper article or a popularized book on gene technology may seek help in a dictionary. He/she needs to know what a gene is. Having found an electronic dictionary designed to help laypersons understand texts, he/she searches for the word *gene* and finds the following result:

gene

the basic unit of inheritance which is transmitted from parents to offspring.

Example 5: Imaginary dictionary article helping laypersons understand texts on gene technology.

The article in Example 5 presents the data in a way that is as easy as possible for laypersons to understand. It is factually correct, devoid of technical jargon and should help users to follow texts in which the term *gene* occurs. In contrast, user profiles of experts and semi-experts show that these user groups have factual and linguistic competences different from those of laypersons and therefore have different needs. A semi-expert may have difficulty understanding a highly scientific text on gene technology and wants to know what a gene actually is. He/she consults the same electronic dictionary and selects a set-up for semi-experts wanting help to understand texts. Example 6 shows the result of his/her search.

gene

A gene is a DNA sequence ending an mRNA (protein), tRNA or rRNA.

Example 6: Imaginary dictionary article helping semi-experts understand texts on gene technology.

The definition in Example 6 is quite different from that written for laypersons. It uses the correct scientific terms, which makes it impossible for laypersons to understand. The comprehension-related information costs for laypersons are prohibitive, whereas they are reasonable for experts and semi-experts. The factual and linguistic competences of intended user groups and their needs in different use-situations directly affect dictionary functions and the ways in which lexicographical data should be presented. Moreover, a reasonable level of comprehension-related information costs does not hinge on definitions being short or long; the definitions belong to different text genres. The relationship between dictionary functions, user needs and use-situations may therefore be used as benchmarks in assessing comprehension-related information costs.

7. Textual condensation may increase information costs

The condensation of lexicographical texts or text parts directly impacts on the process that converts data into useful information. The immediate effect of textual condensation is the reduction of the length and the number of characters of natural sentences or sentence parts for the purpose of saving space. A typical example is the use of the tilde, or swung dash, to replace the lemma inside the article text, usually in collocations and phrases. Lexicographers have a range of options to choose from when condensing texts, Wiegand (1996: 139)

offering the following list: shortening, abbreviating, omitting, shifting, substituting, summarizing and embedding. These condensing operations all involve different types of textual reduction and may be used in printed as well as electronic dictionaries; they are generally found in dictionary articles.

Seen from the user's point of view, condensing operations carried out in dictionary articles may be more or less successful. Some condensed texts are easy to decode, whereas others are difficult to understand. In any case, the reduction of natural sentences to shorter, non-natural sentences or fragments affects comprehension-related information costs, as illustrated in Example 7:

hamster ['hamstɛn], <reg. Vb.; hat> umg. emot. neg.
/jmd./ etw. ~ SYN 'etw. horten'; ↗ FELD I.16.2:
Zucker, Mehl, Lebensmittel ~ ❖ ↗ **Hamster**

Example 7: Article from *De Gruyter Wörterbuch Deutsch als Fremdsprache* (2000).

The article in Example 7 clearly shows that the textual condensation carried out has resulted in non-natural text. The lexicographers may very well feel that their condensing operations (omission, abbreviating, and substituting) have been successful, but most users will probably disagree. The dictionary is intended for text production by learners of German as a foreign language and it seems reasonable to conclude that the comprehension-related information costs are high. In order to help users whose native language is not German, the dictionary contains a list of abbreviations used, but to be able to read the articles fairly easily, these non-native speakers have to learn and memorize no less than 144 abbreviations and their meanings. Even native speakers of German may find this arduous work. In dictionaries designed for the productive needs of non-native speakers of a particular language, non-natural texts or text parts in articles should be used sparingly and with due regard to the factual and linguistic competences identified in the user profile.

Dictionaries designed to provide assistance in communication-orientated use-situations such as text production often contain competence or editorial examples. As illustrated in Example 8, these examples show users how words can be used in connection with other words.

Wörterbuch [...]
ein medizinisches, englisches ~ ; ein historisches,
ein-, zweisprachiges ~ ; [...]

Example 8: Article excerpt from *De Gruyter Wörterbuch Deutsch als Fremdsprache* (2000).

Example 8 contains a typical instance of condensed items giving examples of use; the condensing operations carried out are omitting and substituting. The textual condensation has resulted in non-natural partial sentences, e.g. *ein medizinisches, englisches ~* instead of *ein medizinisches Wörterbuch oder ein englisches Wörterbuch*. And can non-native speakers of German really be blamed if,

after having consulted the above article, they produce words such as *ein-Wörterbuch* and *ein ein-Wörterbuch*? As the presentation of such examples in production dictionaries intended for learners of foreign languages burdens users with high information costs, lexicographers should not condense competence examples in this way.

There is no established and generally accepted yardstick for measuring textual condensation in dictionary articles. Nielsen (2002: 606) proposes the use of a *textual condensation ratio* as an indication of the degree of textual condensation carried out. The textual condensation ratio is calculated as the number of condensing operations divided by the number of elementary items in an article. Condensation ratios below 1 indicate that the number of condensing operations performed is smaller than the number of elementary items. In other words, not all elementary items have been condensed. In contrast, condensation ratios above 1 indicate that the number of condensing operations is larger than the number of elementary items. This means that at least one item has been subject to two or more condensing operations. Although this is a rather crude measure, it points to the level of textual condensation in dictionary articles and moreover also indicates whether this has a significant or insignificant effect on comprehension-related information costs.

8. Cross-references and data distribution affect dictionary use

Lexicographical structures may be used to marshal a dictionary's data into fulfilling its function or functions. Two prominent structures are the distribution structure and the cross-reference structure. The former decides where in the dictionary data is placed, for instance some data about a term may be placed in an article and some in an outer text. The cross-reference structure, also known as the *mediostucture*, is a network structure that consists of the set of relations existing between data by way of cross-references. An analysis of these two structures involves access to and understanding of data, which in turn affect information costs.

Cross-references are widely used in dictionaries to help users acquire further information. It is possible to distinguish between different types of cross-references: cross-references within articles, cross-references from one article to another, cross-references from articles to outer texts or vice versa, cross-references within outer texts, cross-references from one outer text to another, and cross-references to places outside the dictionary. Nielsen (1999: 98-99) shows that it may be beneficial to relate these types of cross-references to dictionary functions, summarized as follows:

- cross-references supporting comprehension
- cross-references supporting text production
- cross-references supporting translation
- cross-references supporting knowledge acquisition

Cross-references should not send users on random searches through the word-list or outer texts. Information overload caused by never-ending cross-reference chains may cause users to end up perplexed and unsure of why they have been referred to different places in the dictionary and what the relationships are between the terms and data concerned. In such cases, temporal costs are involved and so are intellectual costs. One way to minimize lexicographical information costs is for lexicographers to use focused cross-references which support the principal function or functions of the dictionary in question. For example, dictionaries designed for translating may include function-related cross-references enabling users to establish links between data supporting the translation of texts. Lexicographers of polyfunctional dictionaries may elect to use cross-references supporting the primary function only.

Cross-references may be used to establish various types of relationships depending on the functions of dictionaries. One type is hierarchical relationships between lemmatized terms, which may help users understand texts or gain knowledge. It is important to remember that cross-references do not themselves create hierarchical relationships, but connect data in different places enabling users to recreate the relationships on the basis of the data and the links. Example 9 contains three articles illustrating this principle.

defendant *n* a party against whom a criminal charge or civil claim is made: compare PLAINTIFF.

litigant *n* a person engaged in a lawsuit.

plaintiff *n* somebody who brings a civil legal action against another in a court of law: compare DEFENDANT.

Example 9: Articles from *Penguin Office Dictionary* (2006).

Two of the articles in Example 9 contain cross-references to each other and help users to establish a relationship between the terms *defendant* and *plaintiff* from the data in the articles. However, the full potential of using cross-references to assist users in understanding conceptual relationships has not been realized. If the article **litigant** had contained cross-references to the other two articles and these in turn had referred to **litigant**, users would have been able to establish the hierarchical relationship between the three terms. The relationship represented by the data in the referential network would show a species–genus relation, with *litigant* as the generic term and *defendant* and *plaintiff* as specific terms (in this case co-hyponyms). This would have saved users the extra time and effort involved before deducing this terminological hierarchy on the basis of other data that would have to be gathered from elsewhere.

Hierarchical relationships may also be established in other ways depending on the distribution structure and dictionary functions. The data in Example 9 were distributed so that the three articles contain different explanatory data

with 'horizontal' cross-references between the specific terms. Another distribution structure may cause the data to be placed differently:

defendant *n* see LITIGANT.

litigant *n* A litigant is a person engaged in a lawsuit.

A civil lawsuit involves two parties, a plaintiff, who brings a civil legal action against another party called a defendant. In a criminal case the prosecution brings a criminal charge against a party called a defendant.

plaintiff *n* see LITIGANT.

Example 10: Alternative arrangement of data in the articles shown in Example 9.

In the articles in Example 10, the distribution structure has resulted in two cross-reference articles and one synopsis article. Dictionaries designed for helping users to understand texts and develop knowledge may benefit from this type of arrangement where all explanatory data have been placed in one article, and where the cross-reference structure is 'vertical' or 'bottom-up', i.e. from the specific terms to the generic term. Users should be able to create the terminological hierarchy between the three terms without spending too much time and effort. Finally, signposts such as suprasegmental features, for instance italics or colour, could have been used to reduce the time spent on finding the terms *plaintiff* and *defendant* in the synopsis article.

Cross-references may be used to establish links between data in articles and in outer texts. The distribution structure may cause some of the data to be placed in, for instance, a subject-field component and not in the articles. Dictionaries designed for understanding and knowledge building will then need cross-references linking the data so that users may establish the relevant relationship. In some cases, explanatory data in subject-field components may supplement data in articles, and in other cases articles contain no explanatory data; the explanatory data are located in the subject-field component giving an introduction to or an overview of the subject-field treated by the dictionary. Example 11 contains articles and a relevant excerpt from a subject-field component illustrating the interrelationship of the distribution and linking of data.

Civil Division *s* → §2

Court of Appeal *s* → §2; §14

Lord Justice of Appeal *s* → §2; §14

Master of the Rolls *s* → §2

§2 Court of Appeal, Civil Division

The civil division of the appeal court ranks second in the English court hierarchy. Court of Appeal, Civil Division has only jurisdiction to hear civil appeal cases from England and Wales appealed from the High Court and county courts. The president of the court is the **Master**

of the Rolls, and the ordinary appeal judges are called **Lords Justices of Appeal**. Appeal cases are normally heard by 3 judges and appeal lies to the House of Lords if permission to appeal is granted by either the Court of Appeal or the House of Lords.

Example 11: Articles and excerpt from subject-field component in *Engelsk-Dansk Juridisk Basisordbog* (1993).

The four articles in Example 11 are shown as they appear in the dictionary. They do not contain any explanatory, factual data but only cross-references to numbered sections in the subject-field component where the necessary data are given. (Note that this dictionary was designed for Danish users and the subject-field component is written in Danish but has, for practical reasons, been translated into English.) Section 2 of the subject-field component explains the term *Court of Appeal, Civil Division* and the relevant judges. Accordingly, the data fulfil a cognition-orientated function. Moreover, the original Danish text suggests possible translations of the boldface terms, thereby fulfilling a communication-orientated function (*the president* and *ordinary appeal judges*).

Cross-references that support understanding and knowledge development need to work optimally, and lexicographers should carefully consider the search process involved. The cross-reference article **Civil Division** directly specifies its destination, i.e. a numbered section in the subject-field component. The purpose of the article is to give information about the lemma, referring users to a place in an outside text where the relevant data are located. As §2 in the subject-field component deals with the concept of the *Court of Appeal, Civil Division* in an overall way, this type of cross-reference may be called direct. The last two articles in Example 11 contain indirect cross-references because their addresses indicate the location of an overall text part within which the actual addresses are located. The purpose of the article **Master of the Rolls** is to explain the lemma to users by referring them to a place in the outer text. The direct cross-reference address is §2, and the indirect address of *Master of the Rolls* is the identical linguistic unit written in boldface in that section. In order to facilitate access the subject-field component is divided into numbered sections with headlines in boldface and all addresses inside these sections are also written in boldface. This reduces the time spent by users searching for sections and terms explained within sections.

Hierarchical relationships do not only involve genus-species relations. Cross-references may also be used to establish links enabling users to create part-whole relations and generic topic-proper name relations. Moreover, cross-references can be used to link data describing non-hierarchical relationships such as sequential relations. Schaefer (1995: 131-132) shows that two types of sequential relationships are relevant: those based on the linear progression of actions, events, etc. in time or space, and those based on thematic connections (see also Nielsen 1999: 104-105). No matter which function cross-references and the distribution of data support, it is imperative that lexicogra-

phers take care to provide the data in the articles or outer texts necessary for users to establish the hierarchical or sequential relationships. The underlying principle should be that users experience a positive informative value every time explicit cross-references are followed so that any information costs are part of the natural flow of the process.

9. The dual function of lexicographical information costs

The concept of lexicographical information costs is applicable in a range of situations. Its main operative feature is that it has dual applicability in a temporal sense: it can be applied to the time before and the time after the completion of dictionaries. *Retrospective application* is mainly descriptive because it refers to situations that occur after the completion and publication of dictionaries and relates to lexicographical tools already in existence. Examples of retrospective application are: dictionary reviews, metalexicographical analyses and academic descriptions of dictionaries. Reviewers may assess dictionaries on the basis of search-related and comprehension-related information costs in addition to other relevant criteria and issues; and metalexigraphers may analyze existing dictionaries and draw attention to positive and negative cost effects.

Prospective application is generally generative as it refers to situations before the completion and publication of dictionaries and relates to the production of future lexicographical tools. Examples of prospective application are: the planning phase and the compilation phase of dictionaries. Practical lexicographers may prepare instructions for the dictionary work designed to keep information costs at a reasonable level in respect of all the intended dictionary functions and user groups. Theoretical lexicographers may develop new principles for dictionary compiling, making it possible to reduce function-related information costs relative to existing levels and propose new standards for measuring information costs. To sum up, the concept of lexicographical information costs applies to diverse situations that permeate the world of dictionaries, whether such costs go unnoticed or disrupt the natural flow of dictionary consultation, analysis, research or assessment.

10. Conclusion

When we analyze, compile, research or use dictionaries, lexicographical information costs are implicitly or explicitly part of what we do. Users of dictionaries often choose their favourite among their dictionaries when they need help to solve specific problems in specific types of use-situations. Their choice normally depends on the ease and convenience they know or believe are associated with a particular dictionary. Users within the relevant user group elect to consult *Encyclopedia Britannica 2007 Children's Encyclopedia* instead of *Encyclopedia Britannica*, because they know or believe that they can find answers to

their questions without incurring unreasonably high information costs. Lexicographers should strive to present the data in their dictionaries in such a way that users feel they find answers to their questions easily and gain useful knowledge by consulting the dictionaries. Owing to their very nature lexicographical information costs cannot be eliminated, but it is appropriate to draw attention to the fact that prudent and proper consideration may result in a reasonable cost level not seriously affecting the use of dictionaries.

In assessing lexicographical information costs, several factors come into play. Whether costs are high or low in a given situation depends on several interrelated factors, such as the user group, the factual and linguistic competences of users, the structure of data in articles and outer texts, the time it takes to locate the data, the degree of the condensation of data, the distribution of data and the links connecting them. A number of yardsticks have been proposed for measuring search-related and comprehension-related information costs in an attempt to answer the question: how easy (or difficult) is it for users to find and understand the data presented? It has thus been possible to lay the foundation for a framework for evaluating lexicographical information costs for the purpose of dictionary making, use, analysis and research.

References

Dictionaries

- De Gruyter Wörterbuch Deutsch als Fremdsprache* 2000 = **Kempcke, Günther (Ed.)**. 2000. Berlin/New York: De Gruyter.
- Dictionary of Banking and Financial Services* 1985 = **Rosenberg, Jerry M.** 1985. New York: John Wiley & Sons.
- Encyclopedia Britannica* 2007 = **Encyclopedia Britannica, Inc.** 2007.
- Encyclopedia Britannica 2007 Children's Encyclopedia* 2007 = **Encyclopedia Britannica, Inc.** 2007.
- Engelsk–Dansk Juridisk Basisordbog* 1993 = **Nielsen, Sandro.** 1993. Copenhagen: Munksgaard.
- English–Danish Accounting Dictionary* 2005–2008 = **Nielsen, Sandro, Lise Mourier and Henning Bergenholtz.** 2005–2008. <http://www.regnskabsordbogen.dk/regn/gbgb/gbregn.aspx> [25 August 2008].
- Oxford Advanced Learner's Dictionary* 2005 = **Wehmeier, Sally (Ed.)**. 2005. Oxford: Oxford University Press.
- Penguin Office Dictionary* 2006 = **Allen, Robert (Ed.)**. 2006. London: Penguin Books.
- Routledge Dictionary of Economics* 2002 = **Rutherford, Donald.** 2002. London/New York: Routledge.

Other literature

- Bergenholtz, Henning and Uwe Kaufmann.** 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Hermes* 18: 91-125.
- Bergenholtz, Henning and Sandro Nielsen.** 2006. Subject-field Components as Integrated Parts of LSP Dictionaries. *Terminology* 12(2): 281-303.

- Bergenholtz, Henning, Sven Tarp and Herbert Ernst Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, Lothar, Hartwig Kalverkämper, Herbert E. Wiegand together with Christian Galinski and Werner Hüllen (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2: 1762-1832.* Berlin/New York: De Gruyter.
- Casson, Mark.** 2000. *Economics of International Business. A New Research Agenda.* Cheltenham: Edward Elgar.
- De Schryver, Gilles-Maurice.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Liu, Kecheng.** 2000. *Semiotics in Information Systems Engineering.* Port Chester, NY: Cambridge University Press.
- Merrel, Floyd.** 2000. Charles Sanders Peirce's Concept of the Sign. Copley, Paul (Ed.). 2000. *Routledge Critical Dictionary of Semiotics and Linguistics: 28-39.* Florence, KY: Routledge.
- Nielsen, Sandro.** 1999. Mediostructures in Bilingual LSP Dictionaries. *Lexicographica. International Annual for Lexicography* 15: 90-113.
- Nielsen, Sandro.** 2002. Textual Condensation in the Articles of *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Wiegand, Herbert E. (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des »de Gruyter Wörterbuchs Deutsch als Fremdsprache«:* 597-608. Tübingen: Max Niemeyer.
- Nielsen, Sandro.** 2006. A Functional Approach to User Guides. *Dictionaries. Journal of the Dictionary Society of North America* 27: 1-20.
- Peirce, Charles Sanders.** 1992. The Architecture of Theories. Houser, N. and C. Kloesel (Eds.). 1992. *The Essential Peirce. Selected Philosophical Writings, Vol. 1: 285-297.* Bloomington/Indianapolis: Indiana University Press. (First published in *The Monist* 1, 1891: 161-176.)
- Schaeder, Burkhard.** 1995. Mediostrukturen in Fachwörterbüchern. *Lexicographica. International Annual for Lexicography* 11: 121-134.
- Tarp, Sven.** 1998. Leksikografien på egne ben. Fordelingsstrukturer og byggedele i et brugerorienteret perspektiv. *Hermes* 21: 212-137.
- Texas Politics.** 2007. *Key Words and Phrases.* http://texaspolitics.laits.utexas.edu/6_keywords.html [25 August 2008].
- Wiegand, Herbert Ernst.** 1996. Textual Condensation in Printed Dictionaries. A Theoretical Draft. *Lexikos* 6: 133-158.
- Wiegand, Herbert Ernst.** 2000a. Wissen, Wissensrepräsentation und Printwörterbücher. Heid, Ulrich et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8-12, 2000:* 15-38. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Wiegand, Herbert Ernst.** 2000b. Über Suchbereiche, Suchzonen und ihre textuellen Strukturen in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. Wiegand, H.E. (Ed.). 2000. *Wörterbücher in der Diskussion IV. Vorträge aus dem Heidelberger Lexikographischen Kolloquium:* 233-301. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica. International Annual for Lexicography* 21: 202-344.

'n Moontlike model vir die bewerking van idiome in tweetalige woordeboeke*

Liezl Potgieter, *Departement Afrikaans en Nederlands, Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika*
(*liezlpotgieter@gmail.com*)

Opsomming: Woordeboeke is belangrike en noodsaaklike hulpmiddels vir vertalers, maar tog het baie vertalers gemengde gevoelens veral oor tweetalige woordeboeke. Terwyl hulle aan die een kant nie werklik daarsonder kan werk nie, bied tweetalige woordeboeke aan die ander kant dikwels vir vertalers min of geen leiding nie. Dit blyk ook dat tweetalige woordeboeke 'n onvoldoende hulpmiddel vir professionele vertalers by die vertaling van idiome is. Die vertaling van idiome behels meer as net die vertaling van enkelwoorde. In hierdie artikel word 'n moontlike model aangebied om die bewerking van idiome in tweetalige woordeboeke te help verbeter.

Sleutelwoorde: IDIOOM, TWEETALIGE WOORDEBOEK, VERTAALEKWIVALWENT, ABSOLUTE EKWIVALENSIE, GEDEELTELIKE EKWIVALENSIE, DIVERGENSIE, ZERO-EKWIVALENSIE, SURROGAATEKWIVALENSIE, REGISTERVERSKILLE, ONVASTHEDE, TAALBOUSELS

Abstract: A possible model for the treatment of idioms in bilingual dictionaries. Dictionaries are important and necessary resources for translators, but translators nevertheless have mixed feelings particularly about bilingual dictionaries. While on the one hand they cannot really work without them, bilingual dictionaries on the other hand often give translators little or no help. It also appears that bilingual dictionaries are an insufficient resource for professional translators when translating idioms. The translation of idioms entails more than just the translation of single words. In this article, a possible model is presented to help improve the treatment of idioms in bilingual dictionaries.

Keywords: IDIOM, BILINGUAL DICTIONARY, TRANSLATION EQUIVALENT, ABSOLUTE EQUIVALENCE, PARTIAL EQUIVALENCE, DIVERGENCE, ZERO-EQUIVALENCE, SURROGATE EQUIVALENCE, DIFFERENCES IN REGISTER, VARIABLE UNITS, LANGUAGE UNITS

* Hierdie artikel is 'n uittreksel uit 'n ongepubliseerde M.A.-skripsie deur Liezl Gouws, *Die bewerking van idiome in tweetalige woordeboeke: 'n probleem vir vertalers?*, wat in April 2006 aan die Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika, voltooi is.

1. Inleidend

In every language, human beings use idioms. In every language, those idioms share a common property. Decompose any idiom into its constituent words, look at the meaning of those words, and there is no way to reconstruct the idiom. In other words, the word-for-word interpretation of any idiom makes no sense whatsoever. Unless you have a dictionary specifically for translating English-language idioms, give up any hope of translating them. (Weiss 2004)

Die vertaling van idioome veroorsaak talle probleme vir vertalers. Die huidige bewerking van idioome in tweetalige woordeboeke dra geensins daartoe by om vertalers te help om die taak van vertaling te vergemaklik nie.

Alhoewel die ideale oplossing vir die vertaler se probleem 'n volledige lys ten volle bewerkte idioome as buitetekse, of andersins die insluiting van idioome as volwaardige hooflemmas in die vertikale ordening van die sentrale lys sal wees, is daar om praktiese redes, maar ook vanuit die leksikografieteorie talle beperkings en randvoorwaardes wat in gedagte gehou moet word.

'n Lys idioome as 'n buitetekse is 'n gerieflike manier om idioome aan te bied en te bewerk, maar dis dikwels problematies. Die alfabetiese lysing van idioome kan 'n baie ingewikkelde saak wees, omdat dit moeilik is om te bepaal hoe die idioom gelys moet word en die woordorde van die idioom dikwels ook afhang van die teks en konteks waarbinne dit gebruik word. Dit alles maak dit vir die leksikograaf moeilik om idioome op 'n suksesvolle en gebruikersvriendelike wyse binne 'n buitetekse aan te bied en te bewerk. Gevolglik moet daar na alternatiewe aanbiedings- en bewerkingsmoontlikhede gesoek word.

In hierdie artikel word daar na 'n moontlike model vir die verbeterde bewerking van idioome in tweetalige woordeboeke gekyk om te verseker dat dit gebruikersvriendeliker vir vertalers is.

2. Aanbieding

Die onvoldoende manier van aanbieding van idioome in tweetalige woordeboeke is een van die redes waarom hierdie woordeboeke vertalers nie genoegsaam help tydens die vertaling van idioome nie.

Tans word idioome en hulle vertaalekwivalente in tweetalige woordeboeke aangebied as deel van die semantiese kommentaar van die lemmas waaronder hulle aangetref word. Gevolglik word hulle gewoonlik tussen die voorbeeldmateriaal gevind. Alhoewel die idioom op sigself 'n leksikale item is wat eintlik lemmastatus behoort te hê, maak dit in huidige woordeboeke slegs deel uit van die mikrostruktuur. Hierdie werkswyse ontken nie net die status van idioome as leksikale items nie, maar impliseer dikwels ook 'n semantiese verband tussen 'n idioom en die voorafgaande hooflemma. Tog is daar talle ander maniere waarop idioome binne die sentrale teks aangebied kan word wat nie net die bewerking gebruikersvriendeliker sal maak nie, maar ook vertalers sal help om

vinnig en maklik by die korrekte idioom en 'n gepaste vertaalekwivalent uit te kom.

Een van die eerste veranderinge wat leksikograwe kan aanbring om die bewerking van idiome te verbeter, is om idiome te verhef tot makrostrukturele bewerkingseenhede.

'n Onderskeid kan getref word tussen 'n makrostruktuur met 'n vertikale ordening van lemmas en een met 'n vertikale en 'n horisontale ordening (Gouws en Prinsloo 2005: 95). Tradisioneel word slegs vertikaal geordende lemmas as hooflemmas gesien, terwyl horisontaal geordende lemmas gewoonlik lemmas met slegs sublemmastatus is.

Een van die hoofredes waarom leksikograwe horisontale ordening gebruik, is om plek te probeer bespaar. Daar kan ook verdere teksverdigting en plekbesparing bewerkstellig word deur die lemmas se gemeenskaplike aanvangselement weg te laat en met 'n plekbesparingsmerker soos byvoorbeeld 'n tilde (~) te vervang. In artikel 1 uit *Groot Woordeboek* (voortaan GW) kan gesien word hoe die horisontale ordening van lemmas by "bacon" en "bacterial" verder verdig word met behulp van die tilde. In gevalle waar leksikograwe by die horisontale ordening teksverdigting gebruik, kan die gebruiker slegs die lemma via die vertikaal geordende hooflemma bereik. Dit is dus 'n omslagtige proses om by die korrekte vertaalekwivalent van 'n lemma soos byvoorbeeld "bacon rind" uit te kom (sien Artikel 1).

bac' on, spek, varkspek; spekvleis, *bring HOME the ~*, die paal haal; die draai kry; *SAVE one's ~*, daar heelhuids van afkom; *STREAKY ~*, streepspekvleis; ~ **beetle**, spektor; ~**er**, spekvark; ~ **rind**, swoer(d); ~ **roll**, spekrolletjie; ~**y**, spekagtig.
Baco' nian, (-s), Baconvolgeling.
bacter' ial, bakteries; ~ **blight**, vlamsiekte (druive); ~ **disease**, bakteriesiekte; ~ **wilt**, bakteriese verwelksiekte, moko.

Artikel 1: "bacon", "Baconian" en "bacterial" (uit GW)

Sulke lemmas wat deur 'n plekbesparingsmerker en 'n lemmadeel verteenwoordig word en wat slegs via die voorafgaande voluitgeskrewe hooflemma en/of via 'n lemmadeel in 'n nis- of nesekste ingangsposisie bereik kan word, staan as sublemmas bekend.

asyn' . Suur vloeistof deur gisting verkry, bestaande uit water en asynsuur;
asynagtig; asynbottel; asyneksrak; asynerig; asynflessie; asynlug;
asynmakery; asynmoer; asynsmaak; asynsout; asynstandertjie; asynsuur;
asynvaatjie.

Artikel 2: "asyn" (uit VAW)

In gevalle waar die horisontaal geordende lemmas nie verdig is nie, maar volledig uitgeskryf word soos in Artikel 2 uit *Verklarende Afrikaanse Woordeboek*

(voortaan VAW), is die voorafgaande vertikaal geordende lemma nie die enigste manier om die horisontaal geordende lemmas te bereik nie, sodat dit vir die woordeboekgebruiker heelwat makliker is om by die betrokke horisontaal geordende lemma uit te kom. In gevalle waar die horisontaal geordende lemmas volledig uitgeskryf word en 'n streng alfabetiese ordening handhaaf, funksioneer hulle dus nie as sublemmas nie, maar as volwaardige hooflemmas.

Die lys idioome as vertikale hooflemmas is problematies, omdat dit moeilik is om te bepaal waar in die sentrale alfabetiese lys die idioome gelys moet word. Tog is dit belangrik dat die idioome nie slegs deel sal uitmaak van die woordeboek se mikrostruktuur en saam met die voorbeeldmateriaal aangebied word nie, maar dat hulle vollediger bewerk sal word. Deur idioome in die woordeboek as horisontaal geordende lemmas in te sluit, word die probleem uitgeskakel wat sal ontstaan as die idioome vertikaal georden sou word, terwyl dit steeds vir die woordeboekgebruiker maklik is om die idioom op te spoor waarvoor 'n vertaalekwivalent gesoek word. Sodoende kan die idioom van 'n bewerking voorsien word en is die vertikaal geordende lemma nie die enigste manier om by die idioom uit te kom nie. Die aanbieding van idioome in GW en *Pharos 1 Afrikaans-Engels Woordeboek* (voortaan PAEW) sal dus soos in Artikel 3 en Artikel 4 daar uitsien.

bacon, spek, varkspek; spekvleis, **bring HOME the bacon**, die paal haal; die draai kry; **SAVE one's bacon**, daar heelhuids van afkom; *STREAKY* ~, streepspekvleis; ~ **beetle**, spektor; ~**er**, spekvark; ~ **rind**, swoer(d); ~ **roll**, spekrolletjie; ~**y**, spekagtig.

Artikel 3: "bacon" (aangepas uit GW)[†]

bacon vark(spek), spekvleis; rookspek, ontbytspek; **bring home the bacon**, (*infnl.*) die broodwinner wees; ~ *and eggs* eiers met spek; **save one's bacon** die situasie (*of jou bas*) red, heelhuids daarvan afkom. ~ **beetle** spektor. ~ **rind** swoerd. ~ **roll** spekrolletjie, spekvleisrol.

Artikel 4: "bacon" (aangepas uit PAEW)

Leksikograawe kan verder gaan en die aanbieding van idioome in tweetalige woordeboeke verder verbeter deur 'n nuwe tipe neslemmatisering te gebruik. Neslemmatisering kom voor wanneer die leksikograaf van 'n streng alfabetiese ordening afwyk en al die lemmas binne een nes plaas, eerder as om nislemmatisering te gebruik en die lemmas volgens 'n streng alfabetiese ordening te rangskik.

In die geval van nislemmatisering word sowel die vertikale as horisontale lemmas en sowel die hoof- as sublemmas streng alfabeties georden en word daar nie afgewyk van dié ordening nie. Die lemmas binne die nis toon ook 'n interne alfabetiese ordening wat die volgende vertikaal geordende lemma voorafgaan. Nislemmatisering sien soos in Artikel 5 daaruit.

glow, (n) gloed, vuur; rooi kleur; (v) gloei, blaak, brand; ~ *with*, gloei van; ~ **bracelet**, hangkaatser.
glow'er, aanstaar, dreigend (boos) aankyk; ~**ing**, aanstaring.
glow: ~**ing**, gloeiend, vurig, blakend; ~ **-lamp**, gloeilamp; ~ **-worm**, glimwurm, ligwurm.

Artikel 5: "glow" (uit GW)

In die geval van neslemmatisering sien dieselfde lemma soos in Artikel 6 daaruit:

glow *n.* gloed; vuur; rooi kleur; skynsel; blosende kleur; *be in a ~, (fig.)* gloei, gloeiend wees. **glow** *ww.* gloei; brand; smeul; skyn; straal; → **GLOWING** *adj.*; ~ *with* ... brand/gloei van ...; *keep on ~ing* nagloei. ~ **lamp** gloeilamp. ~ **-worm** glimwurm(pie), ligkewer; → **FIREFLY**.
glower ...

Artikel 6: "glow" (uit PAEW)

By die lemmas waar daar neslemmatisering gebruik word, kom daar dus as 't ware teksblokke met die verskillende tipes inligting in tot stand, maar vorm dit steeds 'n geheel, of 'n nes. Indien leksikograwe die aanbieding van idiome in tweetalige woordeboeke verder wil verbeter nadat daar aan die idiome sublemmastatus gegee is, kan neslemmatisering gebruik word om teksblokke binne die artikeltrajek te vorm. Teksblokke bring mee dat soortgelyke datatipes saamgegroeper word. Hierdie soeksones kan byvoorbeeld binne die artikel die volgende behels: benewens blokke vir elk van die vertaalekwivalente tesame met hul semantiese subkommentare met genoegsame konteksleiding en/of voorbeeldmateriaal, ook 'n blok vir kollokasies. Binne die artikeltrajek kan dit 'n blok vir sublemmas insluit. Afsonderlike teksblokke vir idiome kan dan ook aan die artikel van die hooflemma geheg word. Deur hierdie teksblokke vir die bewerking van idiome aan te wend, word verseker dat die idiome nie tussen die ander voorbeeldmateriaal aangebied word nie, maar maklik is om raak te sien en te gebruik omdat die idiome nou 'n prominente posisie as makrostruktuuritems kry. Met behulp van 'n kitstoegangstruktuur kan die teksblok met idiome ook verder vir die gebruiker deur middel van struktuurmerkers soos byvoorbeeld **IDM.** of **UITDR.** uitgelig word. Indien teksblokke soos in Artikel 7 en Artikel 8 gebruik sou word, sal dit soos volg daar uitsien:

bac'on, spek, varkspek; spekvleis, streepspekvleis; **STREAKY** ~; ~ **beetle**, spektor; ~**er**, spekvark; ~ **rind**, swoer(d); ~ **roll**, spekrolletjie; ~**y**, spekagtig.
IDM.: *bring HOME the ~*, die paal haal; die draai kry; *SAVE one's ~*, daar heelhuids van afkom.

Artikel 7: "bacon" (aangepas uit GW)

bacon vark(spek), spekvleis; rookspek, ontbytspek; ~ *and eggs* eiers met spek; ~ **beetle** spektor. ~ **rind** swoerd. ~ **roll** spekrolletjie, spekvleisrol.
IDM.: **bring home the bacon**, (*infnl.*) die broodwinner wees; **save one's bacon** die situasie (*of jou bas*) red, heelhuids daarvan afkom.

Artikel 8: "bacon" (aangepas uit PAEW)

Indien al die voorgestelde verbeterings ten opsigte van die aanbieding van idioome nou saamgevoeg sou word, sal die nuwe verbeterde artikel vir die lemma "bacon" soos volg lyk:

bacon, spek, varkspek; spekvleis, streepspekvleis.
STREAKY ~.
 ~ **beetle**, spektor; ~**er**, spekvark; ~ **rind**, swoer(d); ~ **roll**, spekrolletjie; ~**y**, spekagtig.
IDM.: **bring HOME the bacon**, die paal haal; die draai kry; **SAVE one's bacon**, daar heelhuids van afkom.

Artikel 9: "bacon" (aangepas uit GW)

Deur afsonderlike teksblokke vir die koteksinskrywings en die sublemmas te skep, kan die mikrostruktuur van die artikel en daarmee saam ook die aanbieding van idioome genoegsaam verbeter word sodat die woordeboekgebruiker maklik kan sien in watter geval die teks 'n sublemma, 'n kollokasie, 'n voorbeeldsin of 'n idioom is. Omdat die idioome ook as horisontaal geordende sublemmas binne 'n aparte teksblok aangebied word, help dit vertalers om vinnig en maklik die idioom op te spoor waarna hulle soek. Vertalers se soektog na 'n idioom in die doeltaal waarmee 'n brontaalidioom vertaal kan word, sal op dié manier aansienlik vergemaklik word.

3. Vertaalekwivalente

Die aanbieding van idioome in tweetalige woordeboeke is nie die enigste probleem nie. Die vertaalekwivalente wat verskaf word, sowel as die bewerking daarvan is dikwels 'n selfs groter probleem vir vertalers wanneer dit kom by die korrekte vertaling van idioome.

In die volgende gedeelte sal gekyk word na verskeie moontlike oplossings wat betref die probleem van idioome, hul vertaalekwivalente en hul bewerking in tweetalige woordeboeke.

3.1 Ekwivalensie

Gebruikers raadpleeg meestal tweetalige woordeboeke op soek na 'n vertaalekwivalent waarmee 'n spesifieke woord, uitdrukking of idioom in die brontaal in die doeltaal vervang kan word. Vertalers raadpleeg dus woordeboeke om

hulle te help nie net met teksresepsie nie, maar ook met teksbegrip en teksproduksie. Tog is gevind dat daar aan gebruikers weinig konteks- en koteksleiding verskaf word sodat hulle kan weet watter vertaalekwivalent om binne watter konteks of situasie te gebruik. Hierdie tekort aan konteks- en koteksleiding maak dit vir gebruikers baie moeilik om kommunikatiewe ekwivalensie te bereik.

Wanneer vertalers tweetalige woordeboeke tydens die vertaling van idioome gebruik, is dit omdat hulle op soek is na 'n korrekte vertaalekwivalent vir 'n betrokke idioom. Dit is daarom belangrik dat leksikograwe sal aandag gee nie net aan die verskaffing van vertaalekwivalente nie, maar ook aan die voorsiening van genoegsame konteks- en koteksleiding om die vertaler te help om vinnig en maklik die korrekte vertaalekwivalent op te spoor en reg te gebruik. Erkenning van idioome as sublemmas behoort ook tot 'n volwaardige bewerking te lei. Hierdie konteks- en koteksleiding kan aangebied word in die vorm van voorbeeldmateriaal en addisionele inligting (byvoorbeeld glosse).

In gevalle waar daar 'n verhouding van absolute ekwivalensie tussen die brontaalidioom en die doeltaalidioom bestaan, is daar dus sowel semantiese as kommunikatiewe ekwivalensie tussen die bron- en doeltaal sodat die idioom gewoonlik maklik en probleemloos vertaal kan word. Dit is egter by gevalle waar daar nie 'n verhouding van absolute ekwivalensie en kongruensie bestaan nie, dat vertalers probleme ondervind tydens die vertaling van idioome. In die volgende gedeelte sal gekyk word na die wyse waarop hierdie probleem hanteer kan word deur die bewerking van die vertaalekwivalent te verbeter om vertalers genoegsaam te help om die korrekte vertaalekwivalent vir 'n betrokke idioom binne 'n betrokke konteks te kan opspoor.

3.2 Divergensie

Dit gebeur dikwels dat daar vir 'n brontaalidioom meer as een vertaalekwivalent in die doeltaal bestaan, maar dat die onderskeie vertaalekwivalente slegs gedeeltelike sinonieme is of selfs verskillende polisemiese waardes van die brontaalidioom verteenwoordig. By hierdie gevalle is dit veral belangrik dat die leksikograaf die vertaler genoegsaam sal help om te verseker dat die korrekte vertaalekwivalent binne die korrekte konteks gebruik sal word.

In die geval van leksikale divergensie waar daar 'n verhouding van gedeeltelike ekwivalensie tussen die onderskeie vertaalekwivalente bestaan, is dit eerstens belangrik dat daar vir die woordeboekgebruiker aangedui sal word dat die vertaalekwivalente nie absoluut ekwivalent aan mekaar is nie, maar slegs gedeeltelik. Indien die bron- en doeltaalidiome slegs gedeeltelik ekwivalent is as gevolg van byvoorbeeld registerverskille, is dit nodig dat die leksikograaf nie net die verskille aan die gebruiker sal uitwys nie, maar ook die nodige konteks- en/of koteksleiding sal verskaf sodat die gebruiker sal weet watter vertaalekwivalent om binne watter konteks te gebruik. In Artikel 10 kan gesien word hoe 'n gebrek aan konteks- en koteksleiding dit vir die gebruiker baie moeilik maak om te weet watter vertaalekwivalent om binne watter konteks te kies.

... *s.o. is in ~ for s.t.* iem. kom vir iets in aanmerking, iem. het 'n kans op iets ...

Artikel 10: "line" (uit PAEW)

Dit is veral belangrik dat leksikograwe genoegsame leiding aan woordeboekgebruikers sal verskaf in die geval van semantiese divergensie (met ander woorde in gevalle waar die brontaalidoom polisemies is). Soos gesien kan word in Artikel 11, word polisemie in GW en PAEW tans slegs aangedui met behulp van 'n kommapunt. Daar word gevolglik aan die woordeboekgebruiker min of geen addisionele inligting gegee om te help om die korrekte vertaal-ekwivalent vir die betrokke teks of konteks te kies nie.

... *KEEP an ~ (up)on*, dophou; 'n oog hou oor ...

Artikel 11: "eye" (uit GW)

... *WASH one's ~s of something*; die hande in onskuld was; niks met iets te doen wil hê nie ...

Artikel 12: "hand" (uit GW)

Artikel 11 en Artikel 12 is voorbeelde van artikels waar daar semantiese divergensie voorkom. Nie net word daar nie aan die gebruiker uitgewys wat die betekenisverskille tussen die onderskeie vertaal-ekwivalente is nie, maar die gebruiker weet boonop ook nie watter vertaal-ekwivalent om binne watter tipe teks en/of konteks te gebruik nie. Daar word verder nie vir die gebruiker aangedui dat slegs een van die vertaal-ekwivalente 'n idioom is en die ander net 'n betekenisverklaring nie. Dit is dus belangrik dat die leksikograaf aan die woordeboekgebruiker die nodige konteks- en koteksleiding sal verskaf en die gebruiker op die verskillende betekenisonderskeidings sal attent maak. Dit kan gedoen word met behulp van glosse of voorbeeldmateriaal. Deur glosse met aanvullende inligting soos in Artikel 13 en Artikel 14 in te voeg, word die bewerking gebruikersvriendeliker gemaak. Die leksikograaf moet egter konsekwent te werk gaan by die invoeging van glosse. Hulle moet nie net lukraak gebruik word nie. Dit is belangrik dat alle vertaal-ekwivalente waarvan die konteks of betekenis onduidelikheid kan veroorsaak, van glosse voorsien sal word. Slegs in gevalle waar daar geen onduidelikheid bestaan ten opsigte van die konteks of betekenis nie, kan die glosse as konteks- of betekenisleiding weggelaat word (byvoorbeeld by die tweede vertaal-ekwivalent in Artikel 14). Vanuit die oogpunt van vertalers is die invoeging van hierdie addisionele inligting veral belangrik omdat hulle gewoonlik onder tyddruk werk en dus so vinnig moontlik by die korrekte vertaal-ekwivalent moet uitkom. Deur die glosse in te voeg word die vertalers se soektog beperk tot net een woordeboek en hoef hulle nie nog ander bronne ook te raadpleeg om vas te stel binne watter teks en konteks die betrokke vertaal-ekwivalente gebruik moet word nie.

Aangesien vertalers en ander woordeboekgebruikers gewoonlik heel eerste gaan soek na 'n doeltaalidoom waarmee die brontaalidoom vertaal kan word, is dit dus belangrik dat die onderskeie vertaalekwivalente ook sodanig gelys sal word, met ander woorde deur eers die vertaalekwivalente te gee wat wel idiome is en daarna dié wat slegs enkelwoorde of omskrywings is (sien Artikel 13 en Artikel 14).

... *WASH one's ~s of something*; die hande in onskuld was (nie verantwoordelikheid aanvaar nie); niks met iets te doen wil hê nie ...

Artikel 13: "hand" (aangepas uit GW)

... *KEEP an ~ (up)on*, 'n oog hou oor (sorg vir); dophou (dikwels met 'n onvriendelike bybedoeling) ...

Artikel 14: "eye" (aangepas uit GW)

Ook in gevalle waar daar polidivergensie in die artikel voorkom, is dit belangrik dat daar genoegsame leiding aan die gebruiker verskaf sal word om vinnig en maklik die korrekte vertaalekwivalent te kan identifiseer. In Artikel 15 kan gesien word hoe die bewerking van die idioom "break down" met behulp van konteksleiding verbeter kan word en hoeveel makliker dit is om die korrekte vertaalekwivalent in hierdie aangepaste artikel op te spoor.

... ~ *DOWN*, inmeekaarsak, beswyk (neerval of sterf); in trane uitbars (huil); bly steek (bly staan); teëspoed kry (voertuigprobleme); onklaar raak (breek of gaan staan); ontleed (in dele skei of analiseer) ...

Artikel 15: "break" (aangepas uit GW)

3.3 Zero-ekwivalensie en surrogaatekwivalensie

Net soos by sommige enkelwoorde, bestaan daar ook by idiome gevalle van zero-ekwivalensie, met ander woorde waar daar geen vertaalekwivalente in die doeltaal aanwesig is nie. Dit is egter steeds belangrik dat leksikograwe sulke idiome in die woordeboek sal opneem en dan van 'n betekenisomskrywing of 'n surrogaatekwivalent sal voorsien.

Omdat die woordeboekgebruiker uitgaan van die veronderstelling dat indien die brontaalitem 'n idioom is, die doeltaalitem ook 'n idioom sal wees, is dit belangrik dat die leksikograaf dit pertinent aan die woordeboekgebruiker sal uitwys wanneer 'n idioom van 'n surrogaatekwivalent voorsien word (met ander woorde wanneer die vertaalekwivalent nie 'n idioom is nie) sodat die gebruiker sal weet dat die betrokke vertaalekwivalent slegs 'n betekenisomskrywing of 'n enkelwoord is wat dieselfde betekenis as die brontaalidoom

dra. Dit kan gedoen word deur surrogaatekwivalente van 'n eenvoudige struktuurmerker te voorsien soos byvoorbeeld 'n asterisk (*) of iets soortgelyks (♠, ♣ of ♠) soos in Artikel 16 en Artikel 17 getoon.

... *GO to great (all) ~s, *alles in jou vermoë doen ...*

Artikel 16: "length" (aangepas uit GW)

... ~ *of mind* ♠gemoedsrus, ♠gerustheid ...

Artikel 17: "peace" (aangepas uit PAEW)

3.4 Vals vriende (*faux amis*)

Nog 'n moontlike probleem waarvan leksikograwe woordeboekgebruikers bewus moet maak, is die voorkoms van vals vriende (ook bekend as "*faux amis*"). Vals vriende is gevalle waar dit lyk asof daar 'n ekwivalentverhouding tussen twee woorde (of idioome) op grond van sekere ooreenkomste bestaan, terwyl die twee woorde (of idioome) inderwaarheid nie ekwivalent aan mekaar is nie (sien ook Gouws et al. 2004: 797-806 vir meer inligting oor vals vriende). Hayward en Moulin (1984: 190) beskryf vals vriende soos volg:

Confusion arises because word A (which belongs to the foreign language ...) looks or sounds exactly or nearly like word B, which belongs to the ... mother tongue. The user then establishes an unwarranted interlingual equivalence on the basis of this total or partial similarity.

Een so 'n voorbeeld van vals vriende is die idioome "to pepper someone" en "om iemand te peper". Die betekenis van die Engelse idioom "to pepper someone" is om iemand met vrae te bestook, terwyl die Afrikaanse uitdrukking "om iemand te peper" beteken om iemand te slaan (Prinsloo 2004: 274). Dit is belangrik dat die leksikograaf die gebruiker daarop attent sal maak dat "to pepper someone" nie vertaal kan word met "om iemand te peper" nie (sien Artikel 18). Dit kan byvoorbeeld gedoen word met behulp van 'n glos of addisionele nota na die idioom.

... ~ *s.o./s.t. with ... iem./iets met ... bestook (vrae ens.) [nie iemand peper met ... nie] ...*

Artikel 18: "pepper" (aangepas uit PAEW)

Dieselfde geld ook vir die idioom "make out with someone" en die uitdrukking "met iemand uitmaak". Terwyl die Engelse idioom beteken om met iemand te vry, beteken die Afrikaanse uitdrukking om 'n verhouding met iemand te beëindig. Alhoewel die twee uitdrukkings dus eenders lyk, is hulle betekenis

baie verskillend sodat dit weereens belangrik is dat die leksikograaf die woordeboekgebruikers daarvan bewus sal maak (sien Artikel 19).

... ~ *out with s.o., (Am., infml.)* 'n vryery met iemand hê [Let wel: Hierdie idioom kan nie vertaal word as *met iemand uitmaak* nie] ...

Artikel 19: "make" (uit PAEW)

Sodoende kan die leksikograaf die vertaler help om te voorkom dat idiome verkeerd vertaal word as gevolg van vals vriende wat in die doeltaal voorkom.

3.5 Regstreekse vertalings (leenvertalings)

Nog iets waarvoor leksikograwe moet sorg by die bewerking van idiome in tweetalige woordeboeke is om woordeboekgebruikers (en vertalers) daarvan bewus te maak dat, alhoewel sommige idiome dikwels in die spreektaal regstreeks vertaal word, die regstreekse vertaling (ook genoem "leenvertaling" of "calque") nie altyd die korrekte een is nie. Voorbeelde van sulke gevalle is die Engelse idiome "born with a silver spoon in one's mouth" en "blood is thicker than water". Hierdie idiome word dikwels direk vertaal as "met die silwer lepel in die mond gebore wees" en "bloed is dikker as water", terwyl hulle korrekte Afrikaanse vertalings is "met 'n goue lepel (in die mond) gebore wees" en "bloed kruip waar dit nie kan loop nie".

Weereens, net soos in die geval van vals vriende, kan leksikograwe glosse of notas gebruik om die woordeboekgebruiker (en veral die vertaler) attent te maak op die feit dat hierdie idiome dikwels verkeerd vertaal word (sien Artikel 20 en Artikel 21).

... *born with a SILVER* ~ *in one's mouth*, met 'n goue lepel gebore wees [nie 'n *silwer* lepel nie]; 'n gelukskind wees ...

Artikel 20: "spoon" (aangepas uit GW)

... ~ *is thicker than water* bloed kruip waar dit nie kan loop nie [nie *bloed is dikker* as *water* nie] ...

Artikel 21: "blood" (aangepas uit PAEW)

3.6 Register

Die register van idiome en hulle vertaalekwivalente is ook iets wat leksikograwe duidelik aan woordeboekgebruikers moet uitwys. Dit gebeur dikwels dat die brontaalidioom binne een register val, terwyl die doeltaalidioom se regis-

ters nie almal dieselfde is as dié van die brontaalidoom nie. In sulke gevalle is dit belangrik dat leksikograwe etikette sal gebruik om die verskillende registers aan te dui, anders kan dit maklik gebeur dat 'n vertaalekwivalent vir 'n idioom gekies word waarvan die register nie ooreenstem met dié van die brontaalidoom nie.

Woordeboekgebruikers gaan gewoonlik van die veronderstelling uit dat indien die register van 'n woord of idioom nie aangedui is nie, die woord of idioom se register neutraal is. Tog is dit nie altyd die geval nie.

In PAEW word die register van die idioom "spill the beans" aangedui by sy bewerking onder die lemma "bean" (Artikel 22), maar daar is geen aanduiding van register by sy bewerking onder "spill" (Artikel 23) nie. Die woordeboekgebruiker word ook nie ingelig wat die register van die vertaalekwivalente is en of hulle register dieselfde as dié van die brontaalidoom is nie.

... *spill the ~s*, (infml.) die aap uit die mou laat, met die (hele) mandjie patats uitkom ...

Artikel 22: "bean" (uit PAEW)

... ~ *the beans*, die aap uit die mou laat ...

Artikel 23: "spill" (uit PAEW)

Dit is nie net belangrik dat die aanduiding van die register konsekwent sal geskied nie, maar ook dat alle gevalle waar die register van die idioome nie neutraal is nie, aan die woordeboekgebruiker uitgewys sal word met behulp van etikette, glosse of konteks- of koteksleiding.

Dit is meestal net die brontaalitems wat in GW en PAEW gemerk word, terwyl doeltaalitems oor die algemeen ongeëtiketteer gelaat word. In 'n poging om woordeboekgebruikers te help om nie net kommunikatiewe ekwivalensie te bereik nie, maar ook registerekwivalensie, is dit egter belangrik dat ook die register van die doeltaalitems vir die woordeboekgebruiker, veral die vertaler, aangedui sal word. Indien daar van die vertaler verwag word om die doeltteks formeler of informeler as die bronteks te skryf, is dit byvoorbeeld nodig om te weet watter idioome binne die betrokke tekste en/of kontekste aanvaarbaar sal wees. Ook in gevalle waar die register van die bronteks juis in die doeltteks behoue moet bly, is dit belangrik dat daar vir die vertaler aangedui sal word binne watter register die onderskeie vertaalekwivalente gebruik kan word.

... *be at someone's ~ and CALL*, tot iem. se diens wees; iem. se Klaas wees (infml.); altyd vir iem. klaar moet staan ...

Artikel 24: "beck" (aangepas uit GW)

In Artikel 24 kan gesien word hoe die invoeging van registeretikette soos "infml." dit vir vertalers en ander woordeboekgebruikers makliker maak om te

kan onderskei watter doeltaalidiome se register nie ooreenstem met dié van die brontaalidiom nie.

4. Taalbousels

Die laaste aspek waaraan leksikograwe by die bewerking van idioome in tweetalige woordeboeke kan aandag gee, is die hantering van die verskillende tipes taalbousels (opsionele bykomstige taalbousels, opsionele bykomstige negatiewe woorde en alternatiewe bykomstige taalbousels). Dit is belangrik dat leksikograwe 'n stelsel sal hê waarvolgens elk van hierdie onderskeie tipes taalbousels bewerk sal word.

In die geval van opsionele bykomstige taalbousels, soos byvoorbeeld "hele" en "both" in Artikel 25 en Artikel 26 hieronder, werk die gebruik van hakies goed. Dit is belangrik dat leksikograwe alle opsionele bykomstige taalbousels pertinent sal merk sodat woordeboekgebruikers dit nie vir wesenlike dele van die idioom sal aansien nie. Deur die opsionele taalbousels tussen hakies te plaas, kan vertalers wat die betrokke idioome moet vertaal, duidelik sien watter dele van die idioome wesenlike dele en watter dele opsionele taalbousels is.

spill the ~s, ... met die (hele) mandjie patats uitkom ...

Artikel 25: "bean" (uit PAEW)

... *make (both) ~s MEET*, die tering na die nering sit ...

Artikel 26: "end" (uit GW)

Opsionele bykomstige negatiewe woorde is nog 'n tipe ekstra waaraan die leksikograaf by die aangawe van idioome in tweetalige woordeboeke spesiale aandag behoort te skenk (Combrink 1989: 59). Volgens Combrink (1989: 60) is die beste leiding wat 'n tweetalige woordeboek aan sy gebruikers in dié verband kan gee, die aanduiding van die basiese idioomvorm en die bykomstige opsionele element(e) as sodanig. Dit is daarom belangrik dat woordeboeke net die idioom in sy negatief sal aangee in gevalle waar die idioom slegs in die negatief gebruik kan word of waar die negatief een of meer vertalings het wat nie uit die positiewe aangawe afgelei kan word nie (Combrink 1989: 60).

... *it is no ~ of his/hers* dit is nie sy/haar saak nie, dit gaan hom/haar nie aan nie; *s.t. is of no ~ to s.o.* iets is vir iemand van geen belang nie; *s.t. is of ~ to s.o.* iets is vir iemand van belang ...

Artikel 27: "concern" (uit PAEW)

Die aangawe van die negatiewe vorm in Artikel 27 is dus onnodig en kan weggelaat word aangesien die woordeboekgebruiker maklik genoeg die negatief vanuit die positiewe aangawe kan aflei.

Die laaste groep bousels waaraan leksikograwe spesiale aandag moet skenk, is alternatiewe bykomstige taalbousels. Tans word alternatiewe bykomstige taalbousels wel vir die woordeboekgebruiker aangedui, maar die aanduiding is baie inkonsekwent — in sommige gevalle geskied dit met hakies, in ander gevalle met 'n skuinsstreep en in nog ander gevalle met die woorde "of" of "or". Hierdie verskillende aanbiedings van alternatiewe kan baie verwarrend wees vir gebruikers omdat hulle nie noodwendig sal beseef dat dit deurgaans dieselfde tipe bousel is wat bewerk word nie. Dit is daarom belangrik dat die leksikograaf die data op so 'n wyse sal aanbied dat die gebruiker (of vertaler) maklik sal kan aflei wat presies die leksikograaf probeer meedeel en hoe die vertaler die aangebode inligting in die betrokke artikel in sy/haar teks kan of moet gebruik. Dit is dus nodig dat leksikograwe 'n stelsel vir die bewerking van alternatiewe taalbousels sal ontwikkel en dat dit konsekwent dwarsdeur die hele woordeboek toegepas sal word. Aangesien die hakies vir gebruik by opsionele taalbousels aanbeveel is, sou dit wenslik wees indien leksikograwe moontlik hier die skuinsstreep sal gebruik om die verskillende alternatiewe aan te dui (die gebruik van "of" of "or" neem te veel plek op). Die verbeterde bewerkings sal dus soos in Artikel 28 en Artikel 29 daar uitsien.

... *have an ~ on s.t.* 'n oë/die oog op iets hê ...

Artikel 28: "eye" (aangepas uit PAEW)

... *GO to great/all ~s*, alles in jou vermoë doen ...

Artikel 29: "length" (aangepas uit GW)

'n Laaste aspek waaraan leksikograwe by alternatiewe taalbousels aandag moet gee, is die noodsaaklikheid om alternatiewe bousels altyd pertinent aan te dui, anders as by die idioom "to see something in a certain light" in PAEW waar dit so aangegee word dat dit lyk asof die woord "different" 'n wesenlike deel van die idioom is (Artikel 30), terwyl dit eintlik slegs een van verskeie alternatiewe is wat in die idioom gebruik kan word.

... *see s.t. in a different ~* iets in 'n ander lig beskou ...

Artikel 30: "light" (uit PAEW)

Die aanbieding van hierdie (sowel as ander soortgelyke) idioome kan dus heelwat verbeter word as daar nie net een van die moontlikhede genoem word nie

en as die alternatiewe boonop slegs as alternatiewe en nie as wesenlike dele aangedui word nie. Die verbeterde bewerking van hierdie idioom sal dus soos volg lyk:

... *see s.t. in a different / bad / good* ~ iets in 'n ander / slegte / goeie lig beskou ...

Artikel 31: "light" (aangepas uit PAEW)

Op dié manier word die idioom se alternatiewe taalbousels gelys sodat die bewerking van die idioom wesenlik verbeter word. 'n Vertaler wat met die idioom "to see something in a bad light" gekonfronteer word en dit vervolgens moet vertaal, sal nou maklik kan aflei wat die korrekte vertaalekwivalent vir die idioom is.

5. Ten slotte

In die onderstaande artikels kan gesien word hoe die bewerking van idioome verbeter kan word deur van die voorstelle te gebruik wat in hierdie hoofstuk gemaak is. Artikel 32 is die oorspronklike artikel soos dit in GW aangetref word en Artikel 33 is die verbeterde artikel met pyle en annotasies om die verskillende verbeterings uit te wys.

ball², (n) koeël (geweer); bal, bol; kluit; oogappel; muis (van duim); *ADDRESS the ~*, korrelvat (*gholf*); *~ of the FOOT*, kussinkie (muis) v.d. voet; *HAVE the ~ at one's feet*, die spel in hande hê; op die punt staan om te slaag; *KEEP the ~ rolling*, die spel (die saak) aan die gang hou; *NEW ~*, (*cr.*), nuwe bal; *NO ~*, (*cr.*), foutbal; *be ON the ~*, wakker, byderhand (gereed) wees; *PROVISIONAL ~*, (*golf*), voorlopige bal; *PUT the ~ in the other court*, die verantwoordelikheid op jou teenstander skuif; *SET (START) the ~ rolling*, aan die gang sit, die baan open, die bal aan die rol sit, iets op tou sit, die eerste stoot gee, 'n klip aan die rol sit; *~ and SOCKET*, bolskarnier; *the ~ is WITH you*, dis jou beurt; (v) tot 'n bal vorm; bal.

Artikel 32: "ball²" (uit GW)

Soos gesien kan word, bestaan daar heelwat moontlikhede en maniere waarop die huidige bewerking in tweetalige woordeboeke aangepas kan word om gebruikersvriendeliker te wees en boonop vertalers te help om vinniger en makliker by die korrekte vertaalekwivalent vir 'n betrokke idioom binne 'n spesifieke teks of konteks uit te kom.

Dit is egter belangrik dat leksikograwe ook sal aandag gee aan die voor-tekste-afdeling van die woordeboek en dat daar aan gebruikers 'n omvattende en bruikbare toelichtingstekes verskaf sal word wat verduidelik presies hoe die lemmas, sublemmas en ook idioome, in die woordeboek bewerk word en waarvoor die onderskeie struktuurmerkers, hakies, skuinsstrepe, e.s.m. staan. So-

doende kan leksikograwe daarin slaag om vertalers en ander woordeboekgebruikers te help om idioome met groter sukses te vertaal.

Die kitstoegangstruktuur word verbeter deur idioome in 'n aparte teksblok aan te bied en met 'n struktuurmerker te merk.	<p>ball², (n) koeël (geweer); bal, bol; kluit; oogappel; muis (van duim); <i>ADDRESS the ~</i>, korrelvat (<i>gholf</i>); <i>~ of the FOOT</i>, kussinkie (muis) v.d. voet; <i>NEW ball</i>, (<i>cr.</i>), nuwe bal; <i>NO ball</i>, (<i>cr.</i>), foutbal; <i>PROVISIONAL ball</i>, (<i>golf</i>), voorlopige bal; <i>ball and SOCKET</i>, bolskarnier.</p> <p>IDM. <i>HAVE the ~ at one's feet</i>, die spel in hande hê; op die punt staan om te slaag*; <i>KEEP the ball rolling</i>, die saak aan die gang hou, die spel aan die gang hou (<i>lett.</i>); <i>be ON the ball</i> (<i>inf.</i>), op jou kop wees (<i>infml.</i>), (op en) wakker (<i>infml.</i>), byderhand/gered wees*; <i>PUT the ball in the other court</i>, die verantwoordelikheid op jou teenstander skuif; <i>SET / START the ball rolling</i>, aan die gang sit, die baan open, die bal aan die rol sit, iets op tou sit, die eerste stoot gee, 'n klip aan die rol sit; <i>the ball is WITH you</i>, dis jou beurt; (v) tot 'n bal vorm; bal.</p>	Kollokasies en idioome word van mekaar geskei.
In gevalle waar daar slegs surrogaatekwivalente verskaf is, maar daar tog geskikte DT-idiome bestaan, is die idioome ingesluit. DT-idiome is eerste gelys, en surrogaatekwivalente daarna.		'n Onderskeid word tussen letterlike en nieletterlike idioome getref.
Registeretikette is in sowel die BT as die DT verskaf.		Opsionele taalbousels is tussen hakies aangedui.
Idioome is as sublemmas bewerk en die gemeenskaplike element volledig uitgeskryf en nie met 'n tilde vervang nie.		Alternatiewe taalbousels is m.b.v. skuinstrepe van mekaar geskei.
		Surrogaatekwivalente is met 'n asterisk aangedui.

Artikel 33: "ball²" (Aangepas uit GW)

Nota

† In hierdie artikel, sowel as in die verdere aangepaste artikels, is slegs dié aspek verander wat in die voorafgaande gedeelte bespreek is. 'n Volledig verbeterde artikel word aan die einde van die artikel verskaf.

Bronnelys

Woordeboeke

- Du Plessis, M. (Hoofred.). 2005. *Pharos 1, Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary*. Kaapstad: Pharos Woordeboeke.
- Eksteen, L.C. 1997¹⁴. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans*. Kaapstad: Pharos.
- Labuschagne, F.J. en L.C. Eksteen. 1993⁸. *Verklarende Afrikaanse Woordeboek*. Pretoria: J.L van Schaik.
- Schoonees, P.C. (Hoofred.). 1961. *Woordeboek van die Afrikaanse Taal. Vierde deel H–I*. Pretoria: Staatsdrukker.

Ander bronne

- Combrink, J.G.H. 1989. Wesenlike elemente, ekstras en vulsels in die aangawe van idioome in tweetalige woordeboeke. Botha, T.J.R. (Red.). 1989. *Leksikografie en leksikologie*: 55-77. Menlopark: Serva-Uitgewers.

- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H., D.J. Prinsloo en G.-M. de Schryver.** 2004. Friends will be Friends — True or False. Lexicographic Approaches to the Treatment of False Friends. Williams, G. en S. Vessier (Reds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 797-806. Lorient: Université de Bretagne-Sud, Faculté des Lettres et des Sciences Humaines.
- Hayward, T. en A. Moulin.** 1984. False Friends Invigorated. Hartmann, R.R.K. (Red.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 190-198. Tübingen: Max Niemeyer.
- Prinsloo, A.F.** 2004. *Spreekwoorde en waar hulle vandaan kom*. Kaapstad: Pharos.
- Weiss, J.** 2004. *Translation Hints*. <http://www.lyx.org/TranslationHints> [24 Februarie 2005].

Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria*

Piet Swanepoel, *Department of Afrikaans and Theory of Literature,
University of South Africa, Pretoria, Republic of South Africa*
(swaneph@unisa.ac.za)

Abstract: The reviewing of dictionaries is a time-honoured praxis, but reviewers have often been criticized for only describing the design features of the dictionaries under review and not evaluating these features, for narrowing their reviews down to one or two subjectively chosen design features, and for not explicating their evaluation criteria or indicating their validity.

Improving the quality of dictionary reviews is a task that has been delegated to dictionary criticism as a component of metalexigraphy. In this article, it is proposed that a first step in this direction would be to devise for each dictionary type a set of comprehensive, objective, valid, generally accepted, and operationalizable evaluation criteria. To this end, a general framework for the description and evaluation of dictionary evaluation criteria is developed in this article, using parameters from research on dictionary criticism and the usability of websites.

Keywords: DICTIONARY CRITICISM, DICTIONARY EVALUATION, DICTIONARY EVALUATION CRITERIA, DESCRIPTION OF DICTIONARY EVALUATION CRITERIA, EVALUATION OF DICTIONARY EVALUATION CRITERIA

Opsomming: 'n Voorstel vir 'n raamwerk vir die beskrywing en evaluering van woordeboekewalueringskriteria. Die resenseer van woordeboeke is 'n gerespekterde praktyk, maar resensente is al dikwels daarvan beskuldig dat hulle net 'n beskrywing gee van die ontwerpkenmerke van woordeboeke onder bespreking en nie hierdie kenmerke evalueer nie, dat hulle hul resensies beperk tot een of twee subjektief gekose ontwerpkenmerke en dat hulle nie hulle ewalueringskriteria duidelik uiteensit of hulle geldigheid aandui nie.

Die verbetering van die gehalte van woordeboekresensies is 'n taak wat opgedra is aan die woordeboekkritiek as 'n onderdeel van die metaleksikografie. In hierdie artikel word voorgestel dat 'n eerste stap in hierdie rigting sou wees om vir elke woordeboektype 'n stel omvattende, objektiewe, geldige, algemeen aanvaarde en operasionaliseerbare ewalueringskriteria te ontwikkel. Vir hierdie doel word daar in hierdie artikel 'n algemene raamwerk vir die beskrywing en beoordeling van woordeboekewalueringskriteria ontwikkel deur gebruikmaking van parameters van navorsing oor woordeboekkritiek en die bruikbaarheid van webwerwe.

* This article is an extended version of a paper presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the Woordeboek van die Afrikaanse Taal, Stellenbosch, Republic of South Africa, 30 June–3 July 2008. I would like to thank Rufus Gouws, Howard Jackson and Reinhard Hartmann for their comments on the first draft of this article.

Sleutelwoorde: WOORDEBOEKKRITIEK, WOORDEBOEKEVALUERING, WOORDEBOEKEVALUERINGSKRITERIA, BESKRYWING VAN WOORDEBOEKEVALUERINGSKRITERIA, EVALUERING VAN WOORDEBOEKEVALUERINGSKRITERIA

1. Introduction

Legitime und gute Wörterbuchkritik muss den Werken voraus sein, die sie kritisiert. Sie muss geradezu die Umrisse neuer Werke erschaffen. Solche Wörterbuchkritik ist produktiv und selbst ein Teil der kritischen Wörterbuchforschung. (Wiegand 1994: 3)

Ideally seen, dictionary criticism as praxis is the evaluation, negative or positive, of the design features of a dictionary/dictionaries on the basis of one or more lexicographically relevant evaluation criteria (Ripfel 1989: 49 ff.). Although dictionary criticism/evaluation is a time-honoured practice (cf. Hartmann 1996: 241), it has come under some heavy criticism. Dictionary reviews have been described in derogatory terms as 'primitive' (Béjoint 2000: 113), as often being nothing more than 'incidental sniping' at unmotivated aspects of the design of dictionaries (Osselton 1989: 229), and as being no more than 'repetitions of the publisher's own publicity on the dictionary cover' (Bergenholtz and Tarp 1995: 232).

Even more substantial reviews have been criticized. It has been claimed

- that dictionary reviews often only describe the design features of the dictionary/dictionaries under review and seldom evaluate them (cf. Chan and Taylor 2001, and Osselton 1989),
- that reviewers narrow the scope of their reviews down to an explication and evaluation of one or two subjectively chosen design features of a dictionary, choice of lemmas being a favourite one (cf. Landau 1989: 305-306),
- that reviewers do not adequately motivate the choice of specific design features for evaluation, often focusing on what lexicographers themselves would perceive as trivial with regard to the processes of designing a dictionary or the design of the dictionaries themselves (Osselton 1989: 225),
- that reviewers often do not explicate their evaluation criteria, or indicate what their (lexicographical) status/validity are,[†] and
- that if reviewers make evaluative pronouncements on the design features of dictionaries, the criteria are often only implied and it is left to the reader to deduce them — a strategy whereby reviewers circumvent their task of providing a well-defined and well-motivated set of objective criteria for the evaluation of a dictionary (cf., for example, Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Chapman 1977, Osselton 1989, Ripfel 1989, Rossenbeck 1994, and Steiner 1994).

Of course there are exceptions, but the overwhelming verdict therefore seems to be that reviews often lack objectivity, validity and reliability, and thus are in principle unfit to perform two of their major goals/functions: (1) to 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) to assist lexicographers in optimizing the functionality of their dictionaries.

Improving the quality of the praxis of dictionary criticism is a task that has been left to the scientific study of dictionary criticism/evaluation (*Wörterbuchkritik*) — one of the major components of a theory of lexicography. Although it is by no means clear precisely what the goals, methods and theories of this component of dictionary research should be (cf., however, Wiegand 1994), research in the field suggests that the scientific study of dictionary criticism should encompass the following:

- (1) to improve the quality of the evaluation of dictionaries and the advice reviewers have to give their target readers
 - (a) by providing clear and lexicographically relevant definitions of concepts such as 'evaluation/criticism' and 'positive/negative value' (cf. Ripfel 1989, and Wiegand 1994) and by specifying how to operationalize them,
 - (b) by providing reviewers with sets of explicitly formulated, generally acceptable, comprehensive, systematic and operationalizable criteria for the evaluation of dictionaries of all types,
 - (c) by providing generally acceptable methods to conduct such reviews (cf. Chapman 1977, Hartmann 2001, and Jackson 1996 for a discussion of approaches or methods),
 - (d) by providing the evaluation criteria in formats which are usable/functional for different methods for the review/evaluation of dictionaries of all types, and
 - (e) by providing a theory of the review as genre/document type (and specifying the competencies required from reviewers, goals, target readers and the design guidelines (content, structure, style, presentation, etc.) for the various document types in which reviewers do/could report on their evaluation of dictionaries (cf. Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Jackson 1996, Ripfel 1989, and Rossenbeck 1994), and
- (2) to support lexicographers in the design of dictionaries of all types to optimally achieve the information and communication needs of their intended users in different contexts of use.

In this article, the focus falls on goal (1)(b), viz. that the scientific study of dictionary criticism has to provide reviewers with sets of explicitly formulated, valid, generally acceptable, comprehensive, systematic and operationalizable

criteria for the evaluation of dictionaries of all types. This is a goal which has neither been clearly stipulated nor achieved within the field of the scientific study of dictionary criticism. Working bottom-up from corpora of reviews, the main findings of such studies often only reflect the critique mentioned above, viz. that reviews focus on the design features of dictionaries and seldom evaluate them in terms of clearly specified criteria. As a result, such studies provide lists for the evaluation of dictionaries but these in fact consist for the larger part mostly of the design features of dictionaries which receive attention in reviews — the criteria in terms of which they have to be evaluated often receiving much less attention. Evaluation criteria are often only implicitly formulated, the key concepts are not distinctly defined, and it is not clear what status or validity the criteria have. It is seldom explicitly indicated how they are to be operationalized, and for what kind of reviewing method(s) they are to be used (cf., for example, Chapman 1977, Chan and Loong 1999, Chan and Taylor 2001, Jackson 1996, 2001, Jehle 1990, Kister 1992, Nakamoto 1994, Osselton 1989, Steiner 1984, 1994, and Wiegand 1994, 1998, 2002). Furthermore, with a few exceptions (cf., for example, Ripfel 1989), little effort has been made to explicate, analyse and evaluate the sets of evaluation criteria which are used in the reviews. Little thought has gone into such issues as to what the distinguishing features of evaluation criteria are, how best they can be systematized, in what formats they could best be presented for various kinds of reviewing activities, and how and according to which parameters these evaluation criteria are themselves to be evaluated.

Evaluation criteria for dictionaries are, however, not only presented in dictionary reviews — they also feature prominently and can be harvested from numerous other sources such as lexicographically relevant articles, manuals, handbooks, reports, etc. Although there are exceptions, much the same critique noted above applies to the way in which evaluation criteria are treated in these studies. Although some dictionary types, their design features and evaluation criteria have received more attention in the literature than others, one could hypothesize that for most dictionary types there still does not exist a set of explicitly formulated, valid, generally acceptable, comprehensive, systematic and operationalizable evaluation criteria.

1.1 Goals

The main goal of this article is to report on some groundwork towards the development of a lexicographically motivated framework for the description (analysis and classification) and evaluation of dictionary evaluation criteria. Working bottom-up from lexicographically relevant literature on dictionary evaluation criteria, a number of parameters for the description and evaluation of evaluation criteria will be proposed. Given the lack of research on this topic in the field of lexicography, however, literature on the criteria for the evaluation of documents, specifically Renkema's (1996) 3Cs model, and the frame-

work developed by De Jong and Van der Geest (2000) for the description and evaluation of website design heuristics were also consulted. Relying on usability research in the field of web design for this purpose is not arbitrary. Both the functional approach in the design and evaluation of dictionaries and the usability approach in the design and evaluation of websites (and documents in general) focus in their evaluation on the degree to which the design of the product (a dictionary, a website, a printed document) supports the effectiveness, efficiency and satisfaction with which different kinds of users can perform certain functions in various contexts of use.

In this article, owing to limitations of space, the development of the analytical framework for evaluation criteria is restricted to criteria for printed dictionaries. Although most of them also extend to the evaluation of electronic dictionaries, a number of additional criteria come into play in the case of electronic dictionaries, given the specifics of the medium, an increase in design options, and related usability issues (cf., for example, De Schryver 2003 for an analysis of the design features of electronic dictionaries).

2. A framework for the description and evaluation of evaluation criteria

Because of the lack of research on the parameters for a framework for the description and evaluation of dictionary evaluation criteria, a bottom-up approach is used in this section by incorporating the criteria used in Ripfel (1989) — the first of the more comprehensive studies of dictionary criticism (cf., however, also Jehle 1990, and Nakamoto 1994) — with a strongly modified version of the parameters for the analysis of website heuristics proposed by De Jong and Van der Geest (2000).

Ripfel (1989) describes and classifies dictionary evaluation criteria according to the following parameters:

- (1) the source of the evaluation criterion (journalistic reviews, expert reviews),
- (2) the dictionary/dictionary type to which the evaluation criterion pertains,
- (3) the specific design feature(s) of the dictionary type which the criterion addresses, and
- (4) the evaluation (positive vs. negative) attached to the criterion.

De Jong and Van der Geest (2000) cluster their features for the analysis of heuristics under four major headings, adjusted here to apply to dictionary evaluation criteria:

- (1) the information covered by the evaluation criteria,
- (2) the presentation format of the evaluation criteria,
- (3) the validity of the evaluation criteria, and
- (4) the application of the evaluation criteria.

<p>1. Information covered by the evaluation criteria</p> <p>(a) Specificity General Genre-specific Feature-specific</p> <p>(b) Exhaustiveness Exhaustive — Arbitrary</p> <p>2. Presentation format of the evaluation criteria</p> <p>Instructions Questions Requirements Explanations</p> <p>3. Validity of the evaluation criteria</p> <p>(a) Foundations Standards Theory User research Practitioners</p> <p>(b) Novelty value High — Low</p> <p>4. Application of the evaluation criteria</p> <p>(a) Level of expertise required Mechanistic — Expert</p> <p>(b) Outcome possibilities Open — Closed</p> <p>(c) Operationalization Specified — Unspecified</p>

Table 1: Parameters for the analysis and evaluation of dictionary evaluation criteria derived from De Jong and Van der Geest (2000)

By way of introduction, it seems necessary though to distinguish between statements which merely focus content-wise on the design features of dictionaries and those which are evaluation criteria, i.e. which also indicate what value should be attached to a design feature/design features and why. This is illustrated by means of Jackson's (1996) distinction between internal and external evaluation criteria.

2.1 Descriptive versus evaluative statements

Jackson (1996: 5-6) divides his proposed evaluation criteria for monolingual dictionaries into internal and external evaluation criteria:

Internal criteria derive from what a dictionary says about itself, or what the editors claim for the dictionary. External criteria derive from metalexigraphy, taking into account the linguistic requirements for a lexical description, as well as (*sic*) considerations of dictionary design and production.

As an example of internal criteria, Jackson (1996: 5) gives the claim about their defining strategy made by the compilers of the *New Oxford Dictionary of English* (NODE) in their introduction:

Each word has at least one core meaning, to which a number of subsenses may be attached ... Core meanings represent the typical, central uses of the word in question in modern standard English ... The core meaning is the one that represents the most literal sense the word has in ordinary modern usage ...

Jackson (1996: 5) adds to this quotation that 'these are testable statements', and, indeed, one could check a number of lexical entries and determine if their meanings are defined according to this defining strategy. However, in the quotation above, no evaluative statements about the efficacy of the defining strategy are given — the compilers merely describe how the meaning of words are explained and no claims of an evaluative kind are made, for example, that the defining strategy is better with regard to other defining strategies (or those of other dictionaries) in achieving some functional goal, such as full comprehension of the meaning of a lexical item. Clearly, statements about the design features of dictionaries should be distinguished from evaluative claims about the (in)efficacy of these features with regard to some goal or norm of efficacy (e.g. maximal comprehensibility).

Furthermore, if dictionaries deviate from their defining strategy in certain entries (should specific lexical items require it), it would hardly be considered a lexicographical sin; but in as much as it is an oversight, and occurs often in a dictionary, it would rather entail an infringement of the criterion of consistency (in applying design guidelines). What is at stake here then is an infringement of a high-level evaluation criterion (consistency) which holds for all design guidelines of a dictionary and which has its own motivation (cf. the discussion below).

A common trend in the analyses of dictionary reviews (cf., for example, Chan and Taylor 2001) and in the presentations of sets of evaluation criteria is, however, that the evaluation norms/dimensions (e.g. ease of comprehension of definitions in learner's dictionaries) that motivate a design feature (e.g. the use of a controlled/limited vocabulary in definitions) are simply omitted or only implied. Consequently, the mere absence or presence of descriptive features is then evaluated in themselves as either positive or negative without reference to the relevant motivating evaluation norm.

It should also be obvious from the example given above, that internal and external criteria are inadequately differentiated, and can in fact overlap: internal criteria can also be derived from metalexigraphy, or for that matter from linguistic requirements, or considerations of the design and production of a

dictionary (thus making them less useful parameters in terms of which to analyse evaluation criteria). The defining strategy outlined for NODE is actually followed by other dictionaries as well and not in itself seen as a negative practice.

Reviewers also commonly take the pronouncements made by compilers in the front matter of their dictionaries or those on their covers as a starting point for a review and then clearly link their analysis and evaluation to what Jackson would consider as 'external' criteria. A good example is Feinauer's (2007) review of the *Woordeboek van die Afrikaanse Taal Deel XII*.

2.2.1 Information covered by the evaluation criteria

2.2.1.1 Source of the evaluation criterion

Ripfel (1989) distinguishes between evaluation criteria used in journalistic reviews and those used in expert reviews. The genre of review types is expanded to five subtypes in Bergenholtz and Mogensen (1994) to include the following: extensive review for lexicographers, shorter review for lexicographers, extensive review for users (and lexicographers), shorter review for users (and lexicographers) and the book announcement. Given their differences in target readers, goals and length, one could expect different and more or less aspects of the design of a dictionary to be treated in more or less depth, and perhaps also differences in how explicitly reviewers will outline and motivate their evaluation criteria (cf. Bergenholtz and Mogensen 1994: 10, and Ripfel 1989: 31). However, it is not obvious whether different kinds of evaluation criteria are used in the different types of reviews or whether some of these are restricted to a certain type of review, making the source of evaluation criteria a less useful choice for differentiating between them.

Secondly, given that the same evaluation criteria can be used in any of these subtypes of reviews, it does not seem to be possible to differentiate between these evaluation criteria in terms of the subgenre of reviews in which they are used. However, one could assume that the author(s) of a review and/or the publication in which a review appears, and thus also the set of evaluation criteria used in the review, may affect their status. This issue needs more empirical research and until then the genre subtype or source does not seem to be a useful parameter for the description and evaluation of evaluation criteria other than being important for documentation/referencing and control purposes.

2.2.1.2 Scope of the evaluation criterion

Dictionary evaluation criteria differ in their scope. Parameter (2) (dictionary type) and parameter (3) (design feature) in Ripfel's set of parameters, and parameter (1) of De Jong and Van der Geest (2000) (information covered by the evaluation criteria) refer to the scope of an evaluation criterion. Some evalua-

tion criteria may hold for all dictionary types, others may be dictionary-specific; some evaluation criteria may hold for all the design features of dictionaries, others may only focus on one such design feature. A few examples are given in the following paragraphs.

2.2.1.3 Dictionary type

In his analysis of a corpus of bilingual dictionary reviews, Tomaszczyk (1988) indicates that only three (equivalents, directionality, reversibility) of the nine design features evaluated in his reviews (equivalents, directionality, reversibility, alphabetization, retrievability, redundancy, coverage, currency, reliability) pertain to the bilingual dictionary itself. Tomaszczyk (1988: 289) summarizes reviewers' evaluation criteria with regard to equivalents as follows:

The equivalents should be of an insertable kind, i.e. capable of being used in actual texts, and preferably monolexemic ... The equivalents proposed should be carefully selected closest possible ones rather than cross-linguistic (near) synonyms ... Definitions are allowed only with 'equivalentless' lexis ... Even these (i.e. definitions — PS) should be formulated in such a way (i.e. abbreviated) as to be substitutable ... In view of the fact that one-to-one lexical correspondences across languages are rare, use should be made of meaning (sense) discriminations. In a bidirectional dictionary, ... these should be given in the source language and for every discrete meaning in the target language ...

Obviously, it would be a task for research on dictionary typologies and dictionary structures — two major components of the theory of lexicography — to distinguish between the design features which are unique to a certain dictionary type and dictionary features which hold for all dictionaries, and a task for research on criticism to indicate what criteria are needed to evaluate all these features.

2.2.1.4 General evaluation criteria

Some evaluation criteria are applicable to dictionaries of all types. Good candidates for general or high-level criteria are Renkema's 3Cs model in which he proposes three general criteria, viz. correspondence, consistency and correctness. These evaluation criteria cut across all the other design guidelines and evaluation criteria and may have slightly different interpretations as they apply to each of these guidelines and criteria.

The one criterion most often reverted to in dictionary criticism is the criterion of consistency in the application of design guidelines. Dictionaries often deviate from their design guidelines, a common one being the lexical items they select from the general vocabulary for treatment in the dictionary. For example, De Schryver (2005) indicates at length how many dictionaries transgress this criterion by inconsistencies in the selection or omission of lexical items or in the over- or under-treatment of certain lemmas in dictionary arti-

cles. Other reviewers focus on such aspects as deviations in the ordering of lexical items in the wordlist of a dictionary and their microstructural treatment, or deviations in the defining strategy.

Tomaszczyk (1986: 293) links consistency to the *reliability* of the dictionary (which is increased if lexicographers adopt a set of clearly defined principles and adhere to them consistently). However, consistent application of the design guidelines of a dictionary ensures that users, once they get acquainted with the design of the dictionary, will be able to know exactly what information they can expect to be present in the dictionary and where to find it. In general, though, consistency in the application of the design guidelines leads to dictionaries having certain features, viz. 'predictability, calculability, analyzability and controllability' (Gouws and Prinsloo 2005: 9).

Correctness simply entails that whatever information the lexicographer provides to the user must be correct (although one might have to add the rider: 'in as far as the lexicographer could determine'). In contrast to correctness, correspondence may at first sight seem to be a vaguer criterion. However, Rundell (1999) addresses both these criteria in his deliberation on the improvement in the quality of the information provided in monolingual learner's dictionaries and the improvement in the way this information is presented to meet the needs and expectations of the potential users of a dictionary. More specifically he states that the correctness of a dictionary is optimized in as much as 'the description of a language that a dictionary provides, corresponds more closely to "the truth", that is, to reliable empirical evidence regarding the ways in which the language is used'. Obviously, the relevant empirical evidence referred to here is what one deduces from an acceptable corpus of the use of the language the dictionary purports to describe.

Correctness can of course extend beyond the grammatical information provided in a dictionary as it is a criterion that could be applied to all this information. Closer adherence to the correctness criterion is seen as positive; providing incorrect information is evaluated negatively. Rundell (1999: 88) does, however, relativise mere correctness (for example on the basis of a corpus analysis) against 'lexicographical relevance', i.e. the cluster of factors which enables a lexicographer to distinguish between information that is simply 'true' and information that is relevant to a specific dictionary (and type of user). For instance, *utter* can be used intransitively (as attested by examples in a corpus), but this is not relevant in a learner's dictionary.

Rundell (1999: 83) gives the criterion of correspondence a functional interpretation when he links it to the degree in which the presentation of the information provided in a dictionary corresponds 'more closely to what we know about the reference needs and reference skills of the target audience'. Correspondence is therefore the kind of high-level criterion tested in reviews in which reviewers try and determine how the design features of a dictionary do or do not support the information needs of different kinds of dictionary users with different linguistic competencies and reference skills in different contexts of language use (cf., for example, Bogaards 1996).

2.2.1.5 Feature-specific evaluation criteria

With regard to the third parameter, Ripfel (1989) structures her evaluation criteria according to an extensive set of 15 thematic categories which cover a variety of topics, including various aspects of the design features of the front and back matter and the central list of the reviewed dictionaries (cf. Ripfel 1989: 93-97 for a discussion). This practice links with a strategy used by some researchers (cf., for example, Bergenholtz and Mogensen 1994, Chan and Taylor 2001, Chan and Loong 1999, Jackson 1996, 2001, and Tomaszczyk 1986) in which the various evaluation criteria (focusing on only one design feature each) are organized according to the frame structure of dictionaries (or parts thereof): the front and back matter of a dictionary and their integration with the central wordlist, and various aspects of the design of the central wordlist, its macrostructure (choice of lemmas, lemmatization and outer and inner access structures), its microstructure (data categories and their structure), and the mediostructure (cf. Gouws and Prinsloo 2005 for a recent overview).

A good example is Jackson's (1996: 7-11) proposed criteria for the evaluation of monolingual dictionaries as presented in Table 2. His explication of the relevant criteria is also an example of the variety of ways in which evaluation criteria are presented for prospective reviewers and of how vague they in fact may be.

Content category	Evaluation criteria
Range of vocabulary	'A reviewer would need to determine whether, for its size and scope, the dictionary had adequate coverage of up-to-date, technical, international and, if appropriate, regional lexis.'
Word formation	'The judgment to be made is whether the account of word formation enables a user to ascertain the formal (morphological) relations between words.'
Homographs	No evaluation criterion. (Cf., however: 'The criteria for determining what is a headword have important consequences for lexical description as well as for accessibility.')
Sense division	'The issue here relates both to the adequacy of the lexical description and to how straightforward it is for the user to find the desired sense.'
Defining	'Not only does a reviewer need to assess the adequacy of the definitions, but also whether they are stylistically appropriate for the intended user.'
Beyond definition (lexical relations, collocations, connotations, etc.)	No evaluation criterion.
Pronunciation	'There are two issues here: the transcription system, ... and the accent to be represented.'

Grammar	'Learners' dictionaries have aimed for full coverage, which raises the question about how grammatical information is represented for effective access. A reviewer needs to evaluate how much information about the grammatical operation of words is necessary for a dictionary to fulfil its recording function, as against the need not to provide too much unnecessary information for the intended users.'
Usage	'The extent to which dictionaries are consistent in using their range of usage labels and how they apply them are matters for the critic to evaluate.'
Examples	'A number of questions need to be asked by the critic, relating to: the extent of the use of examples, what role they are seen to play in exemplification, where the examples come from (corpus or invented), and how consistently the dictionary's policy on examples is implemented.'
Etymology	'It is a matter of critical evaluation whether the information in the dictionary under review is appropriate to its size, purpose and intended users.'
Special features, e.g. synonym essays, boxed comments, usage notes, etc.	'The question is whether they add to the lexical description and the coverage and usefulness of the dictionary.'

Table 2: Criteria for the evaluation of monolingual dictionaries proposed by Jackson (1996)

It would be the task of research on dictionary typologies, dictionary structures and dictionary use to define a frame structure for each dictionary type. Much of this has already been done, for example, for the monolingual explanatory dictionary. Besides outlining relevant design guidelines and evaluation criteria for the front and back matter and the macrostructure of this dictionary type, much research focuses on the design of the microstructure. A number of design guidelines exist with regard to (1) what information categories should be included in the treatment of lemmas in the articles of general monolingual dictionaries (e.g. orthography, morphology, word class, meaning, idioms, etc.), (2) how this information should be encoded, and (3) how this data should be structured in the article as a whole or within the sections of the article dealing with specific information categories. (See in this regard the 'classical' data matrix provided in Hausmann and Wiegand (1989) and the linguistically motivated one given in Hudson (1988: 311-312).) Design guidelines for a variety of dictionary types and criteria to evaluate them are discussed extensively in Hausmann et al. (1989: 409-1056).

Dictionary-specific frame structures can be used as basis for the systematization of evaluation criteria for each dictionary type. These frame structures can also be used to identify for which aspects of the design of a dictionary type no relevant or inadequate evaluation criteria have been proposed and to meas-

ure the comprehensiveness and systematicity of the proposed evaluation criteria. (For the use of dictionary functions as a basis for organizing the design features of dictionaries and their associated evaluation criteria, see the discussion below.)

2.2.1.6 Exhaustiveness

The parameter of comprehensiveness ties in with De Jong and Van der Geest's (2000) parameter of exhaustiveness, which refers to how well an evaluation criterion or a set of evaluation criteria cover a design feature/features of a dictionary. If one compares, for example, Jackson's (1996) two criteria for dictionary definitions above (adequacy, appropriate style) with McMillan's (1949) (i.e. dictionary definitions should be complete, clear, accurate, consistent, independent, objective, and neutral), it is obvious that the former are not as exhaustive as the latter (cf. Gouws and Prinsloo 2005: 147-148 for an explanation of each of these criteria).

On the other hand, McMillan's (1949) framework for the evaluation of college dictionaries is more exhaustive, but obviously does not cover every aspect of the design of such dictionaries as can be verified against the frame structures proposed for monolingual explanatory dictionaries (cf. references above). McMillan's criteria relate to the following aspects of the design of such dictionaries (cf. Mdee 2004: 370-371):

- (1) The quantity of the information the dictionary offers. This includes the number of entries, meanings, new words (when compared to competitors), synonyms, pronunciation, etymology and the use of subject and usage labels.
- (2) The quality of the information presented in the dictionary. The aspects to examine here are:
 - (a) the accuracy of the information,
 - (b) the completeness of the information,
 - (c) the clearness of the information,
 - (d) the simplicity in presenting the information, i.e. the information should be rendered in such a way that it can easily be deciphered, and
 - (e) the modernity of the dictionary, i.e. the number of current words in the language which has been included in the dictionary.
- (3) The effectiveness of the presentation of the information. This includes the order of arranging entries, the placement of etymology, the ordering of senses, and the presentation of pronunciation. Also important is the typography.

As a dictionary type, learner's dictionaries have been extensively reviewed with regard to most of their relevant design features, such as the lists provided by Bogaards (1996) and Ilson (1999). Dretzke (1997) has compiled a list of thirteen evaluation criteria, mostly derived from design guidelines for the content categories, to aid students and teachers to assess the value of a learner's dic-

tionary. A much more exhaustive list, however, is provided by Chan and Loong (1999) — a fact that underlines the relativity of a concept such as exhaustiveness when it is measured in terms of the sets of evaluation criteria that have been developed.

2.2.1.7 Value

Ripfel (1989) lists both the positive and negative evaluation criteria for each aspect of the design of the reviewed dictionaries (cf. Ripfel 1989: 137 ff.). For example, one of the positive criteria with regard to the quality of the meaning explanations of a dictionary is also the provision of encyclopedic information ('die Bedeutungserläuterungen geben auch Auskunft über die Sache selbst, der das Wort gilt'); a negative criterion/tendency is the omission of some meaning distinctions (cf. Ripfel 1989: 153).

Both these evaluation criteria are, however, problematic in as much as it is not made clear why they are considered to be positive or negative. For example, the omission of meanings has a negative effect on users finding the semantic information they need. The relevant functional evaluation criterion in this case is that for reception purposes, users must be able to easily find the meaning of the word they may look up (cf. Bogaards 1996 and the discussion below). Providing as many meanings as possible for each lemma furthers this goal; providing less, hampers it.

The status of the positive evaluative statement provided as example is also uncertain. It is based on a distinction between encyclopedic meaning and linguistic meaning, which is difficult to uphold. However, given that such a distinction could be made, it is not clearly stated why it would necessarily be a positive feature if a meaning explanation in a dictionary includes encyclopedic information. Unless this motivation is spelt out, it is rather problematic to assess the validity and usefulness of such a criterion.

Whether it is in fact necessary or useful to make a distinction between positive and negative evaluative statements or evaluation criteria is not clear. Any evaluation criterion can implicitly be seen as being positive (in as much as it is motivated; cf. the discussion below) and any deviation from it as negative to a certain degree. However, as will be discussed below, further empirical research is necessary on the various forms in which evaluation criteria are or could be formulated before this distinction is simply abandoned.

2.2.2 The presentation format of dictionary evaluation criteria

As should be evident from the examples provided so far, evaluation criteria may be formulated in a variety of ways, only some of which are indicated in De Jong and Van der Geest (2000) (cf. Table 1 above): instructions, questions, requirements or explanations. Evaluation criteria are often only implicitly stated and first have to be made explicit in any one of the various formats above.

Some evaluation criteria may be formulated in any one of these formats and a specific format may be necessitated by the function for which the criteria are to be used. For example, McMillan's (1949) criteria for definitions may be formulated as an instruction/requirement ('Definitions must be/have to be complete, clear, accurate, consistent, independent, objective, and neutral.') or as a question ('Are the definitions complete, clear, accurate, consistent, independent, objective, and neutral?').

In this regard, De Jong and Van der Geest (2000) differentiate between process-orientated heuristics and product-orientated heuristics. In lexicographical terms, such a distinction would hold when evaluation criteria are used as design guidelines in the process of compiling a dictionary, typically formulated as instructions in a plan of the dictionary (cf. Gouws and Prinsloo 2005: 9-19), or as criteria for evaluating the dictionary as a finished product, typically formulated as a set of questions in a checklist. A good example is the checklist provided by Chan and Loong (1999) for learner's dictionaries. Little research has been forthcoming, however, on what formats would best suit what kind of function for which evaluation criteria used.

2.2.3 The validity of evaluation criteria

Dictionary reviewers have been criticized for the fact that they evaluate dictionaries in terms of subjective considerations and seldom motivate their choice of evaluation criteria. When evaluating a dictionary, it is therefore important to know how valid the criteria are on which such an evaluation is based. Ripfel (1989) also touches on the importance of distinguishing between valid evaluative statements and those merely representing the personal tastes of a reviewer. As will be indicated below, it is necessary to distinguish between the motivation/validation for a design feature, the criterion that targets it and the value (positive or negative, be it on a scale) that one should associate with such a criterion.

In the rest of this section, the motivation/validation for dictionary evaluation criteria are discussed in more detail.

2.2.3.1 Dictionary research

Given that dictionary criticism as a field of dictionary research is dependent on the other components of lexicographical research, the design features of dictionaries and their associated evaluation criteria may find their validity in or be motivated by any of the other areas of dictionary research. These areas include

- research on the history of dictionaries (which could explain/motivate why dictionaries have certain features in response to their changing social-cultural contexts),
- research on dictionary typologies and the distinctive features of dictionary types,

- research on dictionary use and users such as the linguistic competencies of target users, their reference skills, and their needs in different contexts of use (cf. the functional validity of evaluation criteria discussed below), and
- research on dictionary structure.

As has already been indicated in the discussion above, each component of the theory of lexicography feeds into the description, evaluation and development of dictionary evaluation criteria in various ways. However, with regard to the validity of evaluation criteria, the functional motivation of the design of dictionaries is currently strongly in focus.

2.2.3.1.1 Functional motivation

In recent years, the functional approach has dominated in the field of dictionary design and evaluation (cf. Bergenholtz and Tarp 2003, Nesi 1999, Swanepoel 2001, Tarp 2007, and Wiegand 1998). According to this approach, each element of the design of dictionaries should be determined/motivated by (1) the linguistic competencies and reference skills of the target users of these dictionaries, and (2) the functions dictionaries have to fulfil in various contexts of use (cf. cognitive and communicative functions discussed in Bergenholtz and Tarp 2003, and Tarp 2007), within the confines of their typological features, and subject to the real world constraints under which they are produced.

In as much as the design features of a dictionary successfully supports the target users and their linguistic needs in such use contexts, they are evaluated positively; in as much as they do not, they are evaluated negatively. Degrees of functionality/usability on the positive–negative scale is thus defined in terms of the performance levels achieved by users for certain tasks in which a dictionary with specific design features is used. As such, it accords with the user approach to the evaluation of a dictionary.

This approach differs from that of the expert checklist in which the functionality/usability of a dictionary is defined in terms of the absence (negatively evaluated) or presence (positively evaluated) of certain design features. As Dillon (2001) points out, however, such an approach assumes that functionality can be an inherent feature of a dictionary if it has certain design features. However, for any combination of target users, specific tasks and usage contexts, there could always be certain aspects of the design of a dictionary which might not function optimally. Furthermore, as will be indicated below in a discussion of Bogaards's (1996) evaluation of a set of learner's dictionaries, experts often have to add that they cannot make a final evaluation of the functionality of a design feature of a dictionary before its success (or not) has been empirically tested with users.

A good example of the way dictionaries can be evaluated by determining how their design features do/do not to a degree support their target users with

their reference skills and specific linguistic needs in certain contexts of use is Bogaards's (1996) review of four learner's dictionaries (*Collins COBUILD English Dictionary*, *Longman Dictionary of Contemporary English*, *Oxford Advanced Learner's Dictionary of Current English*, and *Cambridge International Dictionary of English*). Bogaards (1996) spells out the reference skills and linguistic competencies one can assume the target users possess and the design features needed to optimize the usability of the dictionary for them in specific contexts of L2 acquisition. For example, for receptive purposes — on which the focus is here — finding the meanings of problematic words from texts in the dictionary (the findability problem) and understanding the explanations provided of their meaning (the comprehension problem) are critical. Both of these require that dictionaries should have a number of design features to optimize the findability of the relevant meaning(s) and the comprehensibility of the explanations provided (functional motivation). Furthermore, Bogaards explicates what data categories for lexical items have to be provided in a dictionary for a user to acquire a lexical item fully for both receptive and productive uses of such items (linguistic motivation/lexical acquisition).

Bogaards (1996) discusses in more detail what design strategies the various dictionaries use to address the findability and the comprehensibility problem and, in most cases, to what extent they can or could do so successfully or not, or at least to what degree, relative to each other or pending empirical testing with users. The relevant reception design strategies/features, plus the relevant analytical questions Bogaards (1996: 280) provides are summarized in the following table (Bogaards 1996: 315):

Findability	<ol style="list-style-type: none"> 1. Number of meanings explained (How many words and expressions are entered?) <ul style="list-style-type: none"> – morphological tools (to decipher words not in the dictionary) 2. Accessibility of forms (Are all word forms easily accessible?) <ul style="list-style-type: none"> – meaning-related forms – irregular forms 3. Accessibility of multiword expressions (Where can expressions be found?) 4. Structure of entries (What is done to guide the user in longer entries?) <ul style="list-style-type: none"> – use of labels – guiding principles – general layout
Comprehensibility	<ol style="list-style-type: none"> 5. Definitions (How comprehensible are the definitions given?) <ul style="list-style-type: none"> – defining vocabulary – precision – defining style 6. Illustrations, etc. (What types of illustrations or other devices are used to make meanings clear?) <ul style="list-style-type: none"> – pictures

	<ul style="list-style-type: none"> - synonyms, etc. - notes, etc.
	7. Examples (How comprehensible are the examples given?)

Table 3: Reception design strategies/features and analytical questions used by dictionaries, according to Bogaards (1996)

As the relevant evaluation criteria are not spelt out, such a list, however, does not differ from those that merely indicate what aspects of the design of a dictionary should be described in its evaluation. One cannot simply assume, as is often done in usability studies, that the mere presence of these design features would ensure or optimize findability or comprehensibility.

Sometimes the relevant evaluation criteria are easy to formulate. For example, with regard to the ordering of multiword expressions, Bogaards (1996: 286) notes that users with different mother tongues have different search strategies (motivation), and seeing that learner's dictionaries are aimed at an international market (motivation), 'they should ideally mention all multiword expressions in the entries of all relevant content words'. Bogaards (1996) is, however, also well aware of the fact that it is no simple task to come up with explicit and clearly formulated evaluation criteria with regard to all the relevant design features which are supposed to address the findability and comprehensibility problem. For example, with regard to parameter 1 in Table 3, the argument (and derived evaluation criterion) is that if more words are selected for treatment in a dictionary and if more meanings of words are explained, the chances are greater that the user will find the relevant meaning of a word. Bogaards (1996: 282) phrases the criterion as follows: 'As far as *findability for receptive* purposes is concerned, it is clear that the more lexical units there are in a dictionary, the better are the chances that a learner will find what he needs.' More would therefore be evaluated as positive, fewer as less positive.

In this case, however, what constitutes more and fewer is not stipulated independently or in absolute terms — more and fewer are relativised to the number of lemmas treated in the other dictionaries of the same type which are subject to the review. As Bogaards (1996: 282) notes, the interpretation of whether more and fewer should simply be seen in positive/negative terms may not be the correct approach: 'It is difficult to say whether or to what extent the selection of lexical units by one dictionary is better adapted to the needs of the learners than the choice made by another one.' Application of this criterion is also no easy task as, given the differences in design of various dictionaries, it is rather difficult to create a uniformly defined measurement instrument to count treated lemmas.

The following can serve as a second example. If one assumes that target users know the alphabet, the findability of the word form would be optimized if all forms are treated as lemmas, each with its own entries, and a strict vertical alphabetical ordering of all the lemmas are followed. In as much as the macro-ordering design features of a dictionary thus match this strategy, the more

positive it should be evaluated. Bogaards (1996: 284) spells it out: 'The simplest way of finding a word is to look it up in the alphabetical list. Whenever a form is not in its exact alphabetical place, the learner will have to step back from his text and wonder where he might find what he is looking for.' Nesting and niching strategies, including all multiword expressions in an outer text, and complex systems of cross-referencing worsen the findability problem and should therefore be viewed less positively. They are, however, strategies which most printed dictionaries have to follow given space constraints, and thus are not so easy to simply evaluate negatively.

In most cases, an expert reviewer can only indicate which of these design strategies seem to be or could be effective in addressing these two problems, but independent empirical testing with the target groups is necessary to back up the claims reviewers might make about the efficacy of such design features. With regard to the various definition styles used, Bogaards (1996: 293) clearly states:

In spite of the criticisms formulated, it is difficult to say anything definitive about which type of definition is most profitable for L2 learners. Very little research has been done on this topic. Indeed, whatever research there has been done leads to the conclusion that the dictionary does not seem to contribute in a significant way to a better understanding of a written text.

2.2.3.2 Standards

Standards issued by some authoritative organization would represent, ideally, the generally agreed upon and compulsory criteria for the design and evaluation of dictionaries. One such standard is the revised ISO 1951, which specifies a number of general guidelines for the design of dictionaries and specifically for the content categories and their structuring (cf. Le Meur and Derouin 2005). Such design guidelines could, for example, be used to assess to what extent dictionaries of a specific type in fact include the specified content categories and adhere to other prescribed design guidelines. However, adhering to prescribed guidelines in standards does not *per se* guarantee the efficacy of a dictionary. For example, adhering to the guidelines for the structuring of the data may, but will not necessarily, help users actually find the data they require. Whether or not it facilitates the findability of the data will also depend on whether or not the guidelines themselves have been derived from user research on findability and the structuring of information in dictionaries.

One of the major goals of most standards is to foster uniformity and thus compatibility. However, improving the quality of the product — here, all types of dictionaries — may also be a major goal, but then one which could only be attained if the standards as design guidelines or as evaluation criteria are well motivated by research on product/dictionary use and product design/dictionary structure.

2.2.3.3 Theoretical motivation

Evaluation criteria may also be motivated by lexicographical or any other relevant theory, for example, learning and information-processing theory or theories of language production and comprehension (cf., for example, Van de Poel and Swanepoel 2003). Linguistic theories have by and large dominated as motivation for the content and structuring of the grammatical information in dictionaries. Structuralism has played a major role in this regard (cf., for example, Zgusta 1971), but numerous other theories have had an influence on both the selection and the presentation of grammatical information, such as theories of the mental lexicon, theories of semantic networks, frame semantics and cognitive semantics (cf. Geeraerts 2001 and the articles on 'Homonymy and Polysemy' in the thematic part of *Lexicographica, International Annual for Lexicography*, 17: 1-181, 2001).

One of the major influences to come to the fore with the development of corpora and their use in dictionary compilation is the theory of language learning based on the principle of idiomaticity/phraseology/multiword expressions. This has led to a positive increase in the presentation of usage information for multilexical units in monolingual learner's dictionaries (cf. Béjoint 2000: 209-225; the articles on 'The Corpus Approach to Lexicography' in the thematic part of *Lexicographica, International Annual for Lexicography*, 20: 1-129, 2004 and the articles on 'Idioms in Mono- and Bilingual Lexicography' in the thematic part of *Lexicographica, International Annual for Lexicography*, 19: 1-139, 2003).

Certain design features may in fact have both a theoretical and functional explanation. Bogaards (1996), for example, specifies what (grammatical) information has to be included in the entries of learner's dictionaries for a target user to acquire enough lexical knowledge to use these lexical items in reception and production tasks.

2.2.3.4 Lexicographical practice as motivation

There may also be design features of dictionaries which are simply based on lexicographical practice (practice-based heuristics), many of which are conventional features of dictionaries for which no clear motivation may in fact exist.

2.2.3.5 Novelty as motivation

One of the major features of the so-called 'big five' learner's dictionaries is the innovative way in which they try to address a number of the problems that learners have in finding and comprehending the information they seek (cf. Bogaards 1996, Rundell 1999, and Swanepoel 2000). The novelty value of a design feature (as motivation for its use) may lie in the fact that it addresses a general user problem in a new way. However, novelty in design is not *per se* positive. As Swanepoel (2000, 2001) argues, very few of these design strategies have been tested empirically for their efficacy/functionality.

The foregoing discussion necessitates one to draw a distinction between the motivation for a design feature of a dictionary (e.g. a functional consideration relating to the target user or the context of use) and the evaluation of such a feature, i.e. the degree to which it successfully assists the user with a certain task in a specific content of use. For example, the use of a limited vocabulary in definitions in monolingual learner's dictionaries is motivated by the restricted linguistic competencies of its L2/L3 target users. Its functional goal (on which it has to be evaluated) is to enable better comprehension of the definitions provided. However, whether or not such a controlled vocabulary in fact leads to a clear comprehension of the meaning/sense which such a definition aims to impart is another question and subject to empirical testing with a sample of the target users of the dictionary. For example, such a vocabulary is sometimes too restrictive to explain subtle nuances of meaning of polysemous lexical items or meaning differences between related lexical items (synonyms). Although the question of the motivation for a criterion and its evaluative dimension may in fact overlap, it may be necessary to discern them in other cases.

2.2.4 Application of the evaluation criteria

Parameter 4 of De Jong and Van der Geest (2000) (see Table 1) focuses on the way in which criteria can be used for the evaluation of a dictionary. Subparameter 4(a) would refer to the kind of expertise required to determine whether or not a design feature of a dictionary adheres to the criteria. For example, it can be determined almost mechanistically whether information is provided on the pronunciation of a word, but it might be much more difficult or require greater expertise to determine whether or not the definitions in a dictionary adhere to McMillan's (1949) criteria (cf. the discussion above). Likewise, there may be only one or two possible ways in which information on pronunciation could be provided (a closed set), but numerous ways in which Bogaards's (1996) findability and comprehensibility problems may be addressed.

Subparameter 4(c) refers to whether or not the criterion specifies precisely how (or in what respects) it is to be applied or operationalized to come to an evaluation. Some evaluation criteria may require only some matching operation, others are more difficult to use. For example, how does one determine whether the information in a dictionary is easy to find? What would be expected though, is that all evaluative concepts in terms of which a criterion is formulated will be clearly defined and that it should be clearly indicated how to measure them. This is certainly no easy task. As far as could be ascertained, however, no research has been forthcoming on how reviewers go about evaluating dictionaries in practice.

3. Conclusion

From the discussion above, it seems necessary to distinguish between pure

descriptive statements about the design features of dictionaries and evaluation criteria clearly setting out the basis for the evaluation.

For the analysis of dictionary evaluation criteria a framework consisting of the following parameters is proposed here:

<p>1. Information covered by the evaluation criteria</p> <p>(a) Source</p> <p>(b) Scope of the evaluation criteria Dictionary type General evaluation criteria Feature-specific evaluation criteria</p> <p>(c) Exhaustiveness</p> <p>(d) Value</p> <p>2. Presentation format of the evaluation criteria</p> <p>(a) Formulation of items Instructions Questions Requirements Explanations</p> <p>(b) Process vs. product criteria</p> <p>3. Validity of the evaluation criteria</p> <p>(a) Foundations Dictionary research Functional motivation</p> <p>(b) Standards Theoretical motivation Lexicographical practice as motivation Novelty value as motivation</p> <p>4. Application of the evaluation criteria</p> <p>(a) Level of expertise required Mechanistic — Expert</p> <p>(b) Outcome possibilities Open — Closed</p> <p>(c) Operationalization Specified — Unspecified</p>

Table 4: Parameters for the analysis and evaluation of dictionary evaluation criteria

Although the framework presented in Table 4 is mainly meant for the analysis of dictionary evaluation criteria, one could propose, though, that to be usable, the evaluation criteria themselves will have to meet the following evaluation criteria: be explicitly formulated, valid/motivated, generally acceptable, and the evaluative concepts on which they are based will have to be clearly defined and operationalized — features discussed above.

The use of 'towards' in the title of this article is meant to capture the fact that the proposed framework for the description and evaluation of dictionary evaluation criteria is tentative. It is a first step towards the formulation of such a framework, which itself will no doubt change as the empirical base for this study is expanded and more and more studies on dictionary criticism are included.

Endnote

- † The term *evaluation criterion* refers in reviewing practice to any standard, norm, principle, rule of thumb, heuristic or procedure on the basis of which a design feature of a dictionary is evaluated. The term *design feature* itself refers to the 'building blocks' of dictionaries of various types, in particular the frame features of dictionaries, which includes their front and back matter/texts, the macro-, micro-, medio-, and access structures of the central wordlist, the interaction between these outer texts and the central wordlist, and the smaller elements of which these major structural components of printed and electronic dictionaries are constituted.

Bibliography

- Béjoint, H.** 2000. *Modern Lexicography. An Introduction*. Oxford: Oxford University Press.
- Bergenholtz, H. and J.E. Mogensen.** 1994. Wörterbuchkritik in Dänemark. *Lexicographica* 9: 8-35.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes* 31: 171-196.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam: John Benjamins.
- Bogaards, P.** 1996. Dictionaries for Learners of English. *International Journal of Lexicography* 9(4): 277-320.
- Chan, A.C and Y. Loong.** 1999. Establishing Criteria for Evaluating a Learner's Dictionary. Berry, R., B. Asker, K. Hyland and M. Lam (Eds.). 1999. *Language Analysis, Description and Pedagogy*: 298-307. Hong Kong: The Hong Kong University of Science and Technology.
- Chan, A.Y.W. and A. Taylor.** 2001. Evaluating Learner Dictionaries: What the Reviews Say. *International Journal of Lexicography* 14(3): 163-180.
- Chapman, R.L.** 1977. Dictionary Reviews and Reviewing: 1900-1975. Raymond, J.C. and I.W. Russell (Eds.). 1977. *James B. McMillan: Essays in Linguistics by his Friends and Colleagues*: 143-161. Alabama: The University of Alabama Press.
- De Jong, M. and T. van der Geest.** 2000. Characterizing Web Heuristics. *Technical Communication* 47(3): 312-325.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- 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.

- Dillon, A.** 2001. Evaluation of Software Usability. Karwowski, W. (Ed.). 2001. *International Encyclopedia of Ergonomics and Human Factors*: 1110-1112. London/New York: Taylor and Francis.
- Dretzke, B.** 1997. Moderne englisch-englische Lernerwörterbücher: Beurteilungskriterien. *Fremdsprachenunterricht* 41(3): 218-223.
- Feinauer, I.** 2007. Doen die *Woordeboek van die Afrikaanse Taal Deel XII* dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie? *Lexikos* 17: 258-277.
- Geeraerts, D.** 2001. The Definitional Practice of Dictionaries and the Cognitive Semantic Conception of Polysemy. *Lexicographica* 17: 6-21.
- Gouws, H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hartmann, R.R.K.** 1996. Lexicography as an Applied Linguistic Discipline. Hartmann, R.R.K. (Ed.). 1996. *Solving Language Problems*: 230-244. Exeter: University of Exeter Press.
- Hartmann, R.R.K.** 2001. Methods in Dictionary Research. Hartmann, R.R.K. (Ed.). 2003. *Lexicography. Critical Concepts. Volume 3*: 429-444. London/New York: Routledge.
- Hartmann, R.R.K. (Ed.).** 1999. *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Sub-project 9: Dictionaries*. Berlin Free University. <http://web.fu-berlin.de/elc/tnp1/SP9dossier.doc> [3 May 2008].
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. (Eds.). 1989-1991: 328-359.
- 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: Walter de Gruyter.
- Hudson, R.** 1988. The Linguistic Foundations for Lexical Research and Dictionary Design. *International Journal of Lexicography* 1(4): 287-312.
- Ilsou, R.** 1999. Nine Learners' Dictionaries. *International Journal of Lexicography* 12(3): 223-237.
- Jackson, H.** 1996. *Dictionary Criticism*. Unpublished manuscript. Birmingham City University, Faculty of Computing and Information Studies, Research Papers 2-8.
- Jackson, H.** 2002. *Lexicography: An Introduction*. London/New York: Routledge.
- Jehle, G.** 1990. *Das englische und französische Lernerwörterbücher in der Rezension. Theorie und Praxis der Wörterbuchkritik*. *Lexicographica. Series Maior* 30. Tübingen: Max Niemeyer.
- Kister, K.** 1992. *Kister's Best Dictionaries for Adults and Young People: A Comparative Guide*. Phoenix: Oryx.
- Landau, S.I.** 1989. *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Le Meur, A. and M.-J. Derouin.** 2005. ISO 195: A Revised Standard for Lexicography. *Kernerman Dictionary News* 14. <http://kdictionary.com/kdn/kdn14/kdn1401-iso1951.html> [30 April 2008].
- McMillan, J.B.** 1949. Five College Dictionaries. *College English* 10(4): 214-221.
- Mdee, J.** 2004. *Kirkeby's English-Swahili Dictionary*. *Lexikos* 14: 369-395.
- Nakamoto, K.** 1994. *Establishing Criteria for Dictionary Criticism: A Checklist for Reviewers of Monolingual English Learner's Dictionaries*. Unpublished M.A. Thesis. Exeter: University of Exeter.
- Nesi, H.** 1999. The Specification of Dictionary Reference Skills in Higher Education. Hartmann, R.R.K. (Ed.). 1999: 53-67.

- Osselton, N.E.** 1989. The History of Academic Dictionary Criticism with Reference to Major Dictionaries. Hausmann, F.J. et al. (Eds.). 1989: 225-230.
- Renkema, J.** 1996. Over smaak valt goed te twisten. *Taalbeheersing* 18(4): 324-338.
- Ripfel, M.** 1989. *Wörterbuchkritik. Eine empirische Analyse von Wörterbuchrezensionen*. Lexicographica. Series Maior 29. Tübingen: Max Niemeyer.
- Rossenbeck, K.** 1994. Wörterbuchkritik in Schweden nach 1945: Die zweisprachigen Wörterbücher. *Lexicographica* 9: 88-107.
- Rundell, M.** 1999. Thematic Report 6. Recent Trends in Publishing Monolingual Learner's Dictionaries. Hartmann, R.R.K. (Ed.). 1999: 83-98.
- Steiner, R.** 1984. Guidelines for Reviewers of Bilingual Dictionaries. *Dictionaries* 6: 166-181.
- Steiner, R.** 1994. Reviews of Dictionaries in Learned Journals in the United States. *Lexicographica* 9: 159-173
- Swanepoel, P.** 2000. Providing Lexicographic Support for SL Vocabulary Acquisition: What Kind, under What Conditions, for Whom and Why? 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*: 403-418. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Swanepoel, P.** 2001. Dictionary Quality and Dictionary Design: A Methodology for Improving the Functional Quality of Dictionaries. *Lexikos* 11: 160-190.
- Tarp, S.** 2007. Lexicography in the Information Age. *Lexikos* 17: 170-179.
- Tomaszczyk, J.** 1988. The Bilingual Dictionary under Review. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9–14 September 1986*: 289-297. Tübingen: Francke Verlag.
- Van de Poel, K. and P. Swanepoel.** 2003. Theoretical and Methodological Pluralism in Designing Effective Lexical Support for CALL. *Computer Assisted Language Learning* 16(2–3): 174-212.
- Wiegand, H.E.** 1994. Zur Einführung. *Lexicographica* 9: 1-7.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexicographie. 1. Teilband*. Berlin/New York: Walter de Gruyter
- Wiegand, H.E. (Ed.).** 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II*. Tübingen: Max Niemeyer.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

A Lexicographic Approach to Language Policy and Recommendations for Future Dictionaries

Sven Tarp, *Centre for Lexicography, Aarhus School of Business, University of Aarhus, Aarhus, Denmark (st@asb.dk)*

and

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

Abstract: Language policy prevails at different levels and its formulation typically results in a prescriptive presentation of data. In their dictionaries, lexicographers have to respond to the decisions of language policy makers. In this regard dictionaries can adhere to a strict prescriptive policy by including only the prescribed forms. Dictionaries can also give a descriptive account of language use without making any recommendations or claims of correctness. Thirdly, dictionaries can be proscriptive by recommending certain forms, even if such a recommendation goes against the prescribed forms. This article offers an overview of different levels of language policy and the principles of prescription, description and proscription. Examples are given to illustrate certain lexicographic applications of prescription. It is emphasised that access to relevant data is important to dictionary users. Consequently the lexicographic application of proscription is discussed as a viable alternative to prescription. It is suggested that proscription, in its different possible applications, can lead to a lexicographic presentation that benefits the user and that contributes to the satisfaction of the functions of a given dictionary.

Keywords: ACCESS, COGNITIVE FUNCTION, COMPLEMENTARY PROSCRIPTION, DESCRIPTION, DIRECT PRESCRIPTION, DIRECT PROSCRIPTION, DOMAIN-SPECIFIC LANGUAGE POLICY, INDIRECT PRESCRIPTION, INDIRECT PROSCRIPTION, LANGUAGE POLICY, LEVELS OF LANGUAGE POLICY, LEXICOGRAPHIC FUNCTIONS, NATIONAL LANGUAGE POLICY, PRESCRIPTION, PROSCRIPTION, SINGLE PROSCRIPTION, TERMINOLOGICAL LANGUAGE POLICY, TEXT PRODUCTION, TEXT RECEPTION, VARIANTS

Opsomming: 'n Leksikografiese benadering tot taalbeleid en aanbevelings vir toekomstige woordeboeke. Taalbeleid kom op verskillende vlakke voor en tipieserwys lei die formulering daarvan tot 'n preskriptiewe aanbieding van data. Leksikograwe moet in hulle woordeboeke reageer op die besluite van taalbeleidmakers. In hierdie verband kan woordeboeke bly by 'n streng preskriptiewe beleid deur slegs die goedgekeurde vorme in te sluit. Woordeboeke kan ook 'n deskriptiewe verslag van taalgebruik gee sonder om aanbevelings of korrektheidsaansprake te maak. Derdens kan woordeboeke proskriptief wees deur bepaalde vorme aan te beveel, selfs al is so 'n aanbeveling in stryd met die voorgeskrewe vorme. Hierdie artikel bied 'n oorsig oor verskillende vlakke van taalbeleid en die beginsels van preskripsie, deskripsie en proskripsie. Voorbeelde word gegee om sekere leksikografiese toepassings van preskripsie toe te

lig. Daar word beklemtoon dat toegang tot relevante data belangrik vir woordeboekgebruikers is. Gevolglik word die leksikografiese toepassing van proskripsie bespreek as 'n uitvoerbare alternatief tot preskripsie. Daar word aan die hand gedoen dat proskripsie, in sy verskillende toepassingsmoontlikhede, tot 'n leksikografiese aanbieding kan lei wat die gebruiker bevoordeel en wat bydra om aan die funksies van die betrokke woordeboek te voldoen.

Sleutelwoorde: DESKRIPSIE, DIREKTE PRESKRIPSIE, DIREKTE PROSKRIPSIE, DOMEIN-SPESIFIEKE TAALBELEID, ENKELPROSKRIPSIE, INDIREKTE PRESKRIPSIE, INDIREKTE PROSKRIPSIE, KOGNITIEWE FUNKSIE, KOMPLEMENTÊRE PROSKRIPSIE, LEKSIKOGRAFIESE FUNKSIES, NASIONALE TAALBELEID, PRESKRIPSIE, PROSKRIPSIE, TAALBELEID, TEKSPRODUKSIE, TEKSRESEPSIE, TERMINOLOGIESE TAALBELEID, TOEGANG, VARIANTE, VLAKKE VAN TAALBELEID

1. Introduction

The official use of language, on various levels, is typically regulated by strict or less strict rules, formulated by relevant language bodies. Different types of language policies and different approaches to their implementation prevail on the different levels of policy making. A variety of terms have been introduced to refer to various related aspects (cf. Bergenholtz and Gouws 2006). Important in this regard is the distinction between language policy, as applied to the intralingual and the interlingual level, and the notion of communication policy. In this article, *language policy* is used as an umbrella term, referring to all the different levels of decisions and their implementation, aimed at regulating various aspects of language use, especially with regard to spelling, inflection, pronunciation and the formal recognition of words. Lexicographers have to take cognisance of the different forms of language policy relevant to the dictionaries they are compiling, and decisions need to be made regarding their response to the official decisions of language policy makers. They have to negotiate the best ways to ensure that the functions of their dictionaries can be satisfied and the genuine purpose of the dictionary can be achieved. This implies that although they may obey the rules prescribed by formal bodies in the formulation of their language policies, the user needs and functions of the dictionary may at times demand the inclusion of non-prescribed forms. Working with the distinction between prescription, description and proscription, this article hopes to make a contribution to finding a solution for the problems many lexicographers, also within the multilingual South Africa, experience when having to decide on a lexicographic response to language policy.

2. Language policy

Until now, much has been written about description and prescription in terms of lexicographic works. Different authors have argued in favour of one method

or the other as relevant in dictionary making. In this respect, there is a long tradition of relating language policy at a micro level to lexicography. It is therefore surprising that only very few contributions dealing theoretically with the complex relation between language policy at a macro level and the conception of lexicographic works can be found in the existing literature. Cf. Bergenholtz and Gouws (2006) and Bergenholtz and Tarp (2007), of whom the former (Bergenholtz and Gouws 2006: 14) explicitly declares: 'Every single lexicographical decision has a language policy relevance and therefore, in the end, a political dimension'.

This statement has two dimensions: Firstly, it expresses the idea that the lexicographic decisions at the micro level, in one way or another, are related to the general language policy at the macro level. And, secondly, it indicates that the single decisions may also influence the dictionary users' specific language use. It is the first of these two dimensions that will be explored in this discussion, i.e. the relation existing between general language policy and lexicographic decisions. However, in order to discuss this relation, and for the sake of this article, it is necessary to distinguish between at least three different types of general language policy in terms of their coverage, i.e. national, domain-specific and terminological. The necessity of this distinction is due to the fact that these three types of language policy play different roles and mostly deal with different linguistic phenomena, for which reason they have different consequences for lexicography.

In this article, the term *national language policy* refers to the pretended regulation of use within a given speech community, whether this community is only one among others within a specific country (such as some of the African languages in South Africa), the only community in a specific country (like Denmark) or a cross-border community like the German-speaking population in Germany, Austria and Switzerland. The national language policy is normally laid down by some official or semi-official body such as a national language board or an academy and most often prescribes correct spelling and inflection forms and sometimes also pronunciation. In some language communities it even prescribes the words to be used, whereas there are no known examples of a national language policy prescribing style. The boards or academies responsible for the national language policy may enjoy different degrees of authority and there are even speech communities with competing authorities in terms of language policy.

The term *domain-specific language policy* refers to the language policy laid down by 'intermediate' entities such as companies, ministries, universities, local governments, and various kinds of organisations. This type of language policy may regulate the language, or languages, to be used within their sphere of influence, i.e. inside a company or in external communication. Apart from this, the domain-specific language policy most often regulates the style to be used internally and externally in the entity in question. It is normally subordinated to the national language policy, although it may prescribe specific words

and word forms to be used, which in some cases even goes against the recommendations contained in the national language policy.

Finally, the term *terminological language policy* refers to the regulation of terminology within one or several specialised subject fields. This policy may be decided by a national or regional terminological board or by separate organisations, companies or other entities. In this way, it is sometimes interwoven with the domain-specific language policy and sometimes even with the national language policy. The regulation of terminology normally embraces the selection of the recommended terms and their definition.

As can be seen, the three mentioned types of language policy cover different areas and regulate different aspects of language use although they may overlap to a certain degree. This has to be taken into account when planning and compiling different types of lexicographic works. In this respect, lexicographers working with general dictionaries for communicative and cognitive purposes, in one way or another, have to relate to the national language policy; lexicographers (or terminologists) dealing with specialised dictionaries have to relate their work to the relevant terminological language policy; and lexicographers compiling company, branch and similar dictionaries have to relate to the domain-specific language policy as far as it has relevance for their work. However, just as it is important to relate to the language policy at a macro level, it is also important to determine the character of this relation, which will be discussed in subsequent paragraphs, with examples from general and specialised dictionaries. But before proceeding to this discussion, it is necessary to make a brief incursion into the field of methodology in terms of description, prescription and proscription.

3. About the principles of prescription, description and proscription

Prescription, description and proscription represent different methods of collecting and utilising data from different sources, like corpora, linguistic surveys, text investigations, etc. (cf. Bergenholtz 2003). In the interaction between language policy and lexicography, the principles of prescription, description and proscription play an important role. Therefore a brief discussion of these concepts is needed.

3.1 Prescription

By their very nature language boards or official language bodies are prescriptive. They are prescribing, e.g. by formulating the spelling rules for a given language or by acknowledging, for example, certain loan words as belonging to the standard variety of the language. Prescription is not necessarily the acknowledgement of only a single form. Quite often a language body officially recognises different variants, e.g. orthographic variants of a single word. For Afrikaans, the *Afrikaanse Woordelys en Spelreëls* (*Afrikaans Word List and Spelling*

Rules) (Taalkommissie 2001) is the official publication of the Afrikaans Language Commission, a commission entrusted by the South African Academy for Science and Arts to formulate orthographic rules for Afrikaans. This publication presents the prescriptive decisions of the official language body. In many instances, the Afrikaans Language Commission does not give only one form but makes provision for spelling variants, e.g. the forms *weer eens* x *weereens* and *ver* x *vêr*. When it comes to the writing of compounds in which one of the components is a proper noun no less than five variant forms are officially acknowledged and therefore prescribed, e.g. *Kaapstadstasie*, *Kaapstad-stasie*, *Kaapstad-Stasie*, *Kaapstad stasie*, *Kaapstad Stasie*. Prescription prevails on all the different levels of language policy making and can have a major influence on the language use of the relevant speech community.

Prescription strongly comes to the fore in dictionaries. Within a user-driven lexicographic approach, dictionaries are regarded as utility instruments, compiled for a well-defined target user group with specific lexicographic needs in a specific situation. Knowledgeable dictionary users see their dictionaries as practical tools to assist them in solving real problems. Their dictionary consultation should result in an optimal retrieval of information from the data on offer in the dictionary. These users rely on the dictionary to supply the needed data from which they can retrieve the needed information.

Today it is generally accepted that no dictionary can be everything to everyone. When planning dictionaries, lexicographers need to determine the function(s) of the envisaged dictionary, and every aspect of the dictionary should be planned in terms of these functions.

A brief look at dictionaries through the ages show that many lexicographic products had been compiled to display the data the lexicographer decided on — often without having identified a target user group or having taken cognisance of the needs of these users. Far too often the compilation of dictionaries has given no evidence of a functional approach. The overall impression of users had been that the lexicographer knows what should go into a dictionary, that dictionaries contain the truth, the whole truth and nothing but the truth, and that users should therefore rely on dictionaries as authoritative sources of knowledge. Authority has in many dictionaries been seen as equal to the notion of prescription, i.e. where lexicographers inform users how they should use language. This was already evident in Samuel Johnson's approach in preparing his dictionary of 1755 when he says in *The Plan of a Dictionary of the English Language* (1747) that 'fixing the language' was the main purpose of his dictionary. He describes his prescriptive approach as follows: 'Toleration, adoption and naturalization have run their lengths. Good order and authority are now necessary.'

When taking a prescriptive approach lexicographers impose their point of view on the dictionary and the target users of the dictionary. Many dictionary users actually want this kind of guidance, especially when consulting a dictionary for text production purposes. The dictionary should not give choices,

indicate variants or give regional or colloquial words but should present the users with a single, pure, correct form. Different forms and degrees of prescription can be distinguished (cf. Bergenholtz 2003), and these forms will not be discussed in this article. It suffices to say that prescription, especially a strong prescriptive approach, can be regarded as either presenting a single form (with regard to, for example, orthography, pronunciation, meaning or morphological possibilities) or more than one form as the preferred form(s) of the dictionary without any reference to other words from the non-standardised use. This strong prescription implies that only these forms and words should be allowed as being correct whilst their variants or other words should be prohibited. When following a prescriptive approach the lexicographer places himself in the position of having to judge the language and make a decision regarding the accepted forms. In many instances a prescriptive dictionary will follow the rules laid down by a formal language body or entity, as referred to in the section on language policy. The dictionary will then function as an extension and an instrument at the disposal of this standardisation authority and the prescription of the relevant body is presented in the dictionary. The average dictionary user does not distinguish between different language bodies and academies but regard the dictionary as the embodiment of authority. Just as domain-specific language policy sometimes goes against the prescription of the national language policy, lexicographic prescription does not necessarily always adhere to official rules. It also prevails where lexicographers give a single form and ignore all other variants, but where the forms they give deviate from the official language rules with regard to, for example, orthography, morphology, etc. In such a case, the dictionary is not an extension of the language body but follows its own prescriptive methods.

3.2 Description

Description does not only reflect the decisions of an official body or the implementation of an official language policy on any of the levels discussed in a previous section of this article. Description rather endeavours to give a comprehensive account of actual language use by presenting a variety of forms, whether orthographic, morphological or pronunciation variants, or, for example, words representing dialectal, sociolectal or chronolectal variants. Description avoids classifying occurring forms as either recommended or not permissible. It reflects the spectrum of actual language use.

Although official language bodies primarily function in a prescriptive way, even they do sometimes also reflect the use of non-prescribed forms. The latest edition of the already mentioned *Afrikaanse Woordelys en Spelreëls* has a separate text in which a list of so-called 'Omgangsafrikaans', i.e. colloquial Afrikaans, in which words from especially the informal varieties of Afrikaans are given. They are not prescribed as correct or even presented as permissible but they have been included to represent variants and other words from regis-

ters not formally acknowledged by the prescriptive language body. Important, however, is the fact that they have been included in a separate list — a descriptive list complementing the official prescriptive list.

Description plays a significant role in some dictionaries. A descriptive approach in lexicography sees the lexicographer trying to reflect the actual language use, making provision for different variants but without indicating a recommended form or labelling a given form as not permissible. Bergenholtz (2003) indicates different forms and degrees of description. Depending on the functions of a dictionary, a descriptive approach can either frustrate or please the users. In a dictionary compiled for text reception, a presentation of all the different variants can assist the users effectively. Users consulting a dictionary for text production or translation are often frustrated by a descriptive approach, because they do not find explicit guidance regarding the best or the proper or the most correct form. Having to make choices and being confronted with variants is not what they expect from a dictionary.

Description does not imply that all variants are on the same level of acceptability or have the same usage frequency. In some cases, differences in usage frequency are indicated but the dictionary does not commit itself to an indication of a recommended form. A too strong descriptive approach, where a variety of choices are given without an indication of a recommended form, can be frustrating to users, especially where the user is consulting the dictionary for text production purposes. The comprehensive monolingual Afrikaans dictionary the *Woordeboek van die Afrikaanse Taal* (Schoonees et al. 1951–) follows a descriptive approach and tries to be as consistent as possible by employing description in the presentation of different types of data, including pronunciation. Therefore it gives all the pronunciation possibilities of a given word but in this endeavour the descriptiveness often leads to the inclusion of possible instead of real forms. A strong introspective approach leads to a presentation of all possible pronunciation combinations, and these variants are listed without giving the user advice regarding the most typical forms. The article of the lemma *mikro-ekonomie* (English *micro-economy*) illustrates this problem by giving no less than eighteen pronunciation variants.

Owing to the fact that users see dictionaries as authoritative sources, even description can be interpreted by the average user as a presentation of the correct forms. Wiegand (1986) speaks about the normative force of descriptive dictionaries. Although many users do not distinguish between description, prescription and proscription, discussed in the following section, lexicographers need to take a firm decision regarding the approach to follow in their dictionaries. This decision must take cognisance of a range of implications that the opted-for approach may have on the users and the use of dictionaries and language within the given speech community.

Language bodies should be well aware of the advantages and disadvantages of both prescription and description and lexicographers need to realise that both these approaches have an influence on the success of information retrieval in a specific dictionary.

3.3 Proscription

Bergenholtz (2001 and 2003) discusses the notion of proscription and the different forms and degrees of proscription but he also motivates the use of this term. He indicates that this term originates from the Latin word *proscribere* (to make public). Although the English word *proscribe* is used in the sense 'to forbid', the term *proscription* in lexicography does not refer to the state of being forbidden. In a proscriptive approach, the lexicographer wants to inform the user not only about language use but also about the form recommended by the lexicographer (cf. Bergenholtz 2003: 13). In lexicography, a proscriptive approach sees the lexicographer deviating from the prescriptive way of saying 'this should be done' in favour of saying 'this is recommended'. A proscriptive approach often recommends a single form but it may also give different variants or include different words, accompanied by a clear indication of the form the lexicographer recommends. In some specific cases, it could also lead to the recommendation of two or more forms. This may, for instance, be the case when these forms appear with the same frequency in a corpus or when new words or terms are introduced into a given language and the future will decide whether, for instance, a pure loan word, a transliterated word or a coined word will prevail in the speech community. However, although one or more recommended forms are given it does not imply that they are the only permissible forms. It merely represents the lexicographer's recommendation and does not necessarily have to reflect the decisions of a formal language body.

In his classification of different types of proscription, Bergenholtz (2003: 13) differentiates among others between *exact proscription* (where only one variant is recommended; other variants can be mentioned) and *not exact proscription* (where more than one variant is recommended; other variants can be mentioned). Although he refers to the possibility that other variants can be mentioned he does not distinguish a type of proscription based on the mentioning or not of non-recommended forms. Such a distinction may be useful because it contributes to the contextualization which often has an influence on the choices a user makes for a given utterance. In this article, the dichotomy *single x complementary proscription* refers to this distinction. A procedure of single proscription sees only the recommended form(s) included in the dictionary, whereas the recommended forms are complemented by their non-recommended variants in a procedure of complementary proscription. Complementary proscription may be restricted to the article of the lemma representing the recommended form by including the non-recommended forms as microstructural entries in that article. This is *article-internal complementary proscription*. Another form of complementary proscription, however, occurs when all the non-recommended words are lemmatised with cross-references from the articles of the non-recommended forms to the recommended word. This represents *article-external complementary proscription*. A similar distinction can also be introduced for prescription.

From a language policy perspective, the notion of description does not pose many problems. Language policy is not concerned with proscription. Real problems come to the fore with regard to the notion of prescription. Consequently the discussion in this article will move away from description and will focus on prescription and the resulting problems when this approach is reflected in dictionaries. As a solution, arguments will be presented in favour of implementing a proscriptive approach.

4. Problems in existing approaches of prescription

4.1 An example from the Faroe Islands

The Faroe Islands comprise a small group of islands in the middle of the North Atlantic Ocean with a tiny population of about 50 000 inhabitants. Politically, the Faroe Islands is a kind of Danish semi-colony with an extended degree of autonomy. The language spoken by the islanders is Faroese which is an independent Indo-European language with its roots in the Old Norse that came to the islands with the immigration of Norwegian Vikings starting in the 9th century.

For various historical reasons, the Faroe Islands, just as the nearby Iceland, have a strongly purist tradition with an official language policy which prescribes the words to be used in official communication. The problem, however, is that there is big and growing distance between the officially prescribed vocabulary and the vocabulary used by the ordinary people in their daily life. This is reflected both in text reception where the ordinary people may not understand, or at least fully understand, all the prescribed words, and in text production where they will frequently not know which words to use in official communication, e.g. when they approach the local authorities or apply for a job.

These types of communication problems are exactly those which dictionaries should endeavour to solve by means of the selected data. However, the existing monolingual Faroese dictionaries are not primarily planned to satisfy the real needs of the Faroese people in terms of communication, but to serve as agitation tools for the official language purists. As such, they only include the prescribed vocabulary. This approach, which blindly transfers the principles used to formulate the official language policy to lexicography creates a number of serious problems for the Faroese people when they seek assistance to solve text-reception and text-production problems.

In text reception, they will of course be able to access all the prescribed words and find an explanation of their meaning. But when it comes to the non-prescribed words used in colloquial communication, they will have no dictionary to consult as these words are not allowed and therefore not included in the existing monolingual dictionaries.

In text production, the problem is even worse. As the daily used colloquial

words are not allowed in the dictionaries, the ordinary people will find no help at all when they have problems writing informal texts such as personal letters and various kinds of essays. Neither will they receive any assistance when they do not know which words to use in official communication. Although these words are included in the existing dictionaries, the users who do not know them have no way to access them except reading the dictionary from end to end, a time-consuming procedure which is contrary to the very idea of a dictionary as a reference work and consultation tool.

However, the ordinary people are in most cases more creative than their governors. The close relations with Denmark mean that Danish is taught as the first foreign language in the Faroese education system and that most Faroese have a relatively high proficiency level in Danish. Many Faroese with text-production problems in their mother tongue therefore use this competence to consult a bilingual Danish–Faroese dictionary which will then lead them to the prescribed Faroese word.

The same holds true when they need assistance to produce colloquial texts in their mother tongue. In this case, a bilingual English–Faroese dictionary is frequently used for looking up known English words to find the not-prescribed, but daily used Faroese words. However, the precondition for using this method is a relatively high proficiency level in English which many Faroese people lack.

In both cases, the absence of adequate monolingual Faroese dictionaries implies that in order to produce a correct or an informal, colloquial text in their mother tongue, they will have to make a long detour through a foreign language, consulting a dictionary not conceived for this purpose. The reason for this situation is the uncritical transference of principles and methods used within one sphere of human activity, i.e. the formulation of language policy, to another sphere of human activity, i.e. the making of dictionaries which should always be conceived as utility tools with the genuine purpose of meeting the real information needs of the envisaged target group. In this respect, Tarp (2008: 12) writes:

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.

The Faroese example illustrates why the purist approach to lexicographic products can be considered to be highly problematic. On the one hand, this approach leads to dictionaries that do not satisfy the real needs of the user group, in this case the Faroese people. On the other hand, it is counterproductive even from a purist perspective. The very conception of dictionaries ac-

ording to purely purist principles impedes the access to the prescribed words that may never or only rarely be used in real communication because people will be unable to find them and probably not know of their existence.

4.2 Examples from South Africa

The forms prescribed by official language boards, academies and language bodies are usually elevated to the standard variety of the given language. It is interesting to note the way in which dictionaries respond to the decisions of these bodies and the way in which language bodies respond to the decisions made by lexicographers. Dictionaries either function as extensions of these bodies by adhering rigorously to the prescription or, in a descriptive or prescriptive way, they display forms acknowledged by the official bodies as well as other forms and variants not acknowledged by the official bodies.

The response of language boards to the way in which dictionaries negotiate their prescription is not always evident. The 1984 edition of *Tweetalige Woordeboek/Bilingual Dictionary* (Bosman et al. 1984), a bilingual dictionary with Afrikaans and English as language pair, included lexical items as lemmas not yet acknowledged by the Afrikaans Language Commission. The items had been included on account of their usage frequency in Afrikaans. At a subsequent meeting of the Afrikaans Language Commission, one of the members lodged a formal complaint and a letter was written to the publishing house to express the dismay of the Language Commission. Here the dictionary did not wait for the Afrikaans Language Commission to acknowledge the given words but the dictionary reflected actual language use in a descriptive way.

The development of Afrikaans lexicography gives ample evidence of instances where dictionaries took a prescriptive approach (cf. Gouws and Ponelis 1992 and Gouws 1995). In the multilingual South Africa, language contact is a daily reality. Language contact necessarily leads to all languages occurring in the contact situation to be influenced by the other languages and to include borrowed forms in their lexicon. This has happened in all the South African languages, including South African English, as is clearly seen in *A Dictionary of South African English on Historical Principles* (Silva 1996; cf. also Gouws 1999).

An exaggerated purist approach and a persecution of Anglicisms with the consequent prescription had a detrimental influence on, among others, Afrikaans dictionaries. Bergenholtz and Gouws (2006) discuss this issue and mention that even the eighth edition of *Groot Woordeboek/Major Dictionary* (Eksteen et al. 1993) still included forms like *bruismelk/skummelk/roomsmelk* instead of the frequent *melksommel* (for the English *milkshake*), *knormoer* instead of *selfaansitter* (for the English *self starter*) and *briewebesteller* instead of *posman/posbode* (for the English *postman*). The *Tweetalige Woordeboek/Bilingual Dictionary* omits the highly frequent Afrikaans word *geboortemerk* (English *birthmark*) in favour of the Dutch form *moederlek* whilst *bookmark* gets the equivalents *boeklêer* en *leeswyser* but not the much used word *boekmerk*. These contrived forms were

prescribed in order to avoid the use of loan translations from English. Prescription led to a gap between the dictionary and its users because the dictionary did not reflect the real language use.

The comprehensive multivolume WAT, the *Woordeboek van die Afrikaanse Taal*, exercises its descriptive approach among others by giving lemmas representing lexical items from all the different lects of Afrikaans. The focus is especially strong on dialectal and chronolectal variants, although borrowings also find their way into the macrostructure. Given its comprehensive nature and the limited number of Anglicisms, little criticism has been directed at this dictionary in terms of its deviation from a purist approach. Owing to the traditional prescriptive approach of many dictionaries, users came to expect prescription and a presentation of the pure and the only correct form in their dictionaries, especially dictionaries of a more restricted nature. When a number of frequently used borrowings from English were included in two editions of HAT, the *Verklarende Handwoordeboek van die Afrikaanse Taal* (Odendal and Gouws 2000/2005) and this inclusion was mentioned in the preface, the majority of critical remarks was directed at the inclusion of these because the dictionary was no longer portraying pure language.

It is necessary for lexicographers to focus on the needs of their intended target users and the data these users require to ensure that the functions of the dictionary can be satisfied. New dictionary projects should take cognizance of the way in which older dictionaries, dictionary users and language bodies had responded to this issue and the influence it had on the language, the speech community and the lexicographic practice. A distinction should be made between needs as the users would traditionally interpret them and needs realized by the lexicographers in terms of the functions resulting from the user-profile. It amounts to a distinction between presumed needs and real needs. A purely prescriptive approach is often only directed at the presumed needs whereas description and proscription also account for the real needs.

For dictionaries to be compiled in the African languages of South Africa, important decisions need to be taken regarding the application and influence of a strong prescriptive approach. Nong, De Schryver and Prinsloo (2002) already discussed various issues with regard to the problem of loan words versus indigenous forms. There are frequent discussions regarding the preference given to loan words at the cost of already existing ones and the application of spelling rules and the occasional failure of dictionaries to adhere to these rules. The inclusion of coined words in dictionaries is also contentious, because of the confusion it can cause amongst members of the speech community. A dictionary like the *New Sepedi Dictionary* (Prinsloo and Sathekge 1996) counters these objections by explicitly indicating that it has been compiled on the basis of frequency of use.

Lexicographers are faced with challenges regarding the lemmatisation of science and technology terminology. Many of these terms are regarded as 'book terminology' that never occurs in actual language use. Dictionaries should rather display transliterated loan words instead of unnaturally formed lexical

items. The Sesotho sa Leboa word *phaekukunama*, a lexical form referring to the English *meat pie*, was coined more than three decades ago and was included in the *Northern Sotho Terminology and Orthography No. 4* (Wolff 1988) but it is not used by the speech community. (We thank Ms W.M. Mojabelo for this example.)

A similar prescriptive discrepancy between data presented in dictionaries and the forms used by members of the speech community is evident in the inclusion of lemmas like *matengwa* (English *lottery*), *bofora* (English *fraud*) and *taba* (English *topic*) in the Sepedi dictionary *Pukuntšu Dictionary* (Kriel 1983) instead of the forms *lothari*, *froto* and *thopiki* respectively, loan words used by the average members of the speech community. (We thank Prof. D.J. Prinsloo for these examples.)

The problems referred to do not only regard general language, but also specialised language and terminology. Van der Merwe (2008) mentions a number of terms which are frequently used by ordinary people but not accepted as correct terminology by the South African wine industry. One such example is the English term *drinking wine* (*drinkwyn* in Afrikaans), which in the official wine terminology has been replaced by the term *wine* although many ordinary wine drinkers still use *drinking wine*. Another example is *tapvat*, which is the official Afrikaans equivalent for the English term *boxed wine*. In this case, most ordinary people will use the colloquial *bokswyn* which, however, also refers to a lesser quality wine and is therefore not allowed by the industry as an Afrikaans equivalent for *boxed wine*. A last example is the popular word *champagne*, which the European Union has now restricted only to products coming from the French district of Champagne. The official term to denominate wines produced by the same method, e.g. the Cap Classique, but not coming from the French district is *sparkling wine* although *champagne* is still widely used in colloquial language. Thus, if *drinkwyn*, *bokswyn* and *champagne* are not included in a wine dictionary, at least as cross-reference entries, many users of this dictionary will probably not be able to find the correct terms prescribed by the wine industry.

It is clear from these examples that, similar to the situation in the Faroe Islands, the prescribed words and variants often go against the actual language use. By adhering to a prescriptive approach lexicographers can easily isolate their dictionaries from the needs of their users.

5. Access

As utility instruments dictionaries need to convey the type of data determined by their functions. This data is directed at satisfying the real needs of the intended target user group. Consequently dictionaries should not only be regarded as show cases of language prescription but as utility tools they should include all data necessary to fulfil their specific functions, even if this goes against or beyond the prevailing prescriptive vision. The ideal situation would prevail when the decisions of the language bodies are not isolated from the

language used by the speech community at large. Although lexicographers have to take cognisance of the decisions of language bodies and should not endeavour to position their dictionaries in opposition to these decisions, the principles guiding these decisions may not be a dominating force when selecting data for inclusion in their dictionaries. What is included in a dictionary should assist the users in their daily communication, both in text reception and text production in the formal and less formal registers, as well as in satisfying those cognitive needs that fall within the scope of lexicographic products. In a multilingual society especially, translation is important, and dictionaries need to support their users by also employing a translation function. The title of Komensky's dictionary of 1631, *Ianua Linguarum Reserata* (The Gate of Tongues), portrays the assignment of the lexicographer, i.e. the compilation of a product that unlocks the gate of language for the users. When planning dictionaries, lexicographers should go even further by not only unlocking the gate of tongues but by ensuring access to the data behind the gate.

Giving access to relevant data needs to be regarded as one of the most important tasks of a lexicographer. Access to 'relevant data' is not guaranteed if the lexicographer regards that specific data entry in isolation. Contextualisation of data is paramount and in dictionaries this contextualisation often implies not isolating a word from its variants. Access to a given word can often best be achieved via a variant. In their discussion of a dictionary of gene technology, Bergenholtz, Kaufmann and Tarp (1994) refer to an example where a purely prescriptive approach would have impeded the access of users to specific forms. The Spanish equivalent for the English form *DNA* is *ADN*. However, evidence from a Spanish corpus shows that in 40% of the instances the English form *DNA* is used. Although the Spanish equivalent *ADN* may be both the recommended and the prescribed form successful access to this form will often depend on the inclusion of the English loan form *DNA* as a separate lemma of which the treatment can be restricted to a cross-reference entry, guiding the user to *ADN*. A purely prescriptive approach would have led to the inclusion of only the prescribed form *ADN*. The omission of the loan form *DNA* would have impeded the access of many users to the recommended form. The best possible access to the lemma *ADN* can be achieved by means of article-external complementary proscription procedures. The importance of including both *ADN* and *DNA* in the dictionary can also be motivated on practical grounds. If only one form is included a user, not familiar with the other form, may encounter the form omitted from the dictionary in the literature. While so many different acronyms are being used in the scientific field this user may regard the encountered form as a wholly different acronym not related to the form included in the dictionary.

In order to ensure optimal access to the lemma representing the recommended lexical item, lexicographers should not only look at recommended and recognised forms but also at forms that are not allowed and even forms representing frequently made spelling mistakes. For example, users often consult a dictionary to retrieve the meaning of a given word or its translation equivalent,

not realising that they are misspelling the word. If the specific spelling mistake is a very frequent one, the chances are good that different dictionary users would try to access the treatment of the given word via the incorrect form, cf. the inclusion of the article of the lemma *akceptere* in *Den Danske Ordbog* (Hjorth and Kristensen 2003–2006):

akceptere vb. – *alm. stavfejl for* → acceptere.

[akceptere vb. – *a frequent misspelling of* → acceptere.]

Here the lemma is not presented in bold, like the other lemmas in this dictionary. In this way a distinction is made between the allowed and the non-allowed spelling variants. Omitting such a misspelled form from the dictionary will impede the access of many users to the required data. When endeavouring to achieve optimal access, lexicographers should aim to present as many relevant forms as possible. They should include allowed but non-recommended forms as well as not allowed words and words representing typical spelling mistakes, i.e. not allowed and therefore also not recommended words that still occur frequently in language use. Such an inclusion and the resulting treatment (cf. the remarks made in paragraph 6 of this article) will ensure the satisfaction of the need that prompted the dictionary consultation but it will also see the user getting access to and being informed of the correct spelling of the given word. This additional access possibility may lead to a bonus in the information retrieval process.

The forms recommended in a proscriptive dictionary may be words, e.g. words from the standard or a non-standard variety, but also orthographic or inflectional variants. For some of the South African languages, the relevant language boards focus in their prescription merely on forms from one dialect, eschewing the rest (cf. Mojela 2007). When employing a more comprehensive inclusion approach a dictionary will still give a recommended word but the words from other dialects or registers will also be included as lemmas and their treatment will guide the user to the lemma representing the recommended form. This also applies to orthographic variants where the variants are separated in the alphabetical ordering of lemmas. Access between a recommended and non-recommended form does not only go via the outer access routes, i.e. that part of the access process leading a user to the required lemma. When it comes to orthographical variants where the alphabetical ordering does not separate the forms and, especially, in the case of inflectional variants, the inner access route will guide a user between the different forms. Where there are different inflectional variants, for example, the internal ordering of the variants is important to ensure the required access in order to reach the recommended form but also take cognisance of the other forms. One way of differentiating between the recommended and non-recommended forms is by ordering the recommended before the non-recommended. The use of notes (cf. paragraph 6) will place an additional emphasis on the distinction. On an article-internal level, access to the non-recommended form will then always go via the recommended form.

In spite of the prescription of language bodies, lexicographers should realise that their users need access to relevant data. Even if the lexicographer agrees with the decisions of a prescriptive body, the lexicographic presentation should make allowance for different points of departure and different access routes in the access process of different users, for example by including the prescribed Sepedi form *Hlakola* (English *February*) as well as the frequently used form *Feberware*. Many users may try accessing the dictionary via a non-prescribed form and a strict adherence to the prescriptive principle, informing users that only the given form(s) should be used, would then result in unsuccessful dictionary consultation. The inclusion of as many relevant forms as possible will ensure that access to the recommended form, which might or might not be the prescribed form, can succeed by employing more than one possible access route.

6. Presentation

Once the non-prescribed (not allowed) or non-proscribed (not recommended) words and orthographic and inflectional variants have been included in the lemma list in order to prepare user-friendly access routes, the next question that arises is how to treat and present these words and variants as well as their prescribed or proscribed counterparts. As mentioned above, Bergenholtz (2003) operates with various degrees of prescription and proscription. In the following, this idea will be taken a step further and systematised on the basis of a distinction between the mentioned degrees of prescription and proscription as well as their differentiated treatment in different types of dictionary articles. Such a systematisation is necessary not only in order to develop lexicographic theory in this respect, but also to give useful advice and recommendations to lexicographic practice.

6.1 How to present the non-prescribed and non-proscribed forms

As to the non-prescribed or non-proscribed words and variants which are selected as lemmas in order to provide quick and easy access to their prescribed or proscribed counterparts, there are at least five possible types of treatment which, at the same time, represent various degrees of prescription or proscription, namely:

- (a) Not giving a note addressed to the lemma,
- (b) Giving a note indicating that the variant is not allowed or recommended,
- (c) Giving a note indicating that the variant is frequently used, but not allowed or recommended,
- (d) Giving a note indicating that the variant is not allowed, but recommended, for example because it is frequently used, and

- (e) Giving a note indicating that the variant is not allowed in formal or scientific language, but may be used in colloquial language.

It is possible to make a further subdivision in various directions but for the purpose of this contribution this typology is sufficient. Each type can now be considered separately:

- (a) As the selected variants form part of the access route to the prescribed or proscribed forms, it is presupposed that some sort of cross-reference to these forms is provided in each instance. This means that the first example is a pure cross-reference article which does not have any prescriptive or proscriptive value in itself.

Bergenholtz (2003) divides prescription into open and hidden. In his definition, an open prescription requires that the prescriptive value of the data provided is explained somewhere in the dictionary, for example in the users' guide, whereas a hidden prescription is not explained at all. However, it is also necessary to distinguish between information found in the users' guide (or elsewhere in the outer matter) about the prescriptive, and proscriptive, value of the lexicographic data — among them the cross-references — and the same information retrieved from data provided in the respective dictionary articles. In this article, these two different ways of providing this information will be called *direct* and *indirect* prescription and proscription respectively.

Thus, because many users may not read the users' guide or remember its indications, they will probably in many cases not understand the prescriptive or proscriptive value of the cross-reference, i.e. why they are referred to another lemma. For this reason, it could be recommended to use *direct* prescription and proscription and add more data in order to assist the user, even when the information extracted from this additional data is not relevant for the dictionary function, e.g. text production, in the strict sense of the word.

- (b) In the second type, the data required above are addressed to the lemma. These data can be presented in various ways, for example as it has been done in the Danish–English *Genteknologisk ordbog* (Bergenholtz and Kaufmann 1992) which among its functions includes text production for mother-tongue speakers of Danish:

fermentere <vb> *ferment* <vb>
Fermentere er en ældre betegnelse for at gære
[*Fermentere* is an old term for *gære*]

Here the user is informed that the verb *fermentere* is no longer used in Danish specialised language dealing with molecular biology and has been replaced by *gære* which is actually a common Danish word, also used in colloquial communication. In this way, the user is tactfully and indirectly recommended to use the term *gære* in modern specialised language dealing with molecular biology.

(c) In the third type, the user acquires the additional information that the non-prescribed or non-proscribed form is frequently used. One such example is taken from the Danish monolingual production dictionary *Den Danske Netordbog* (Bergenholtz et al. 2008):

liniere *verbum* <-r, -de, -rt>

Dansk Sprognævn godtager ikke denne skrivemåde, selvom den er mere end 25 gange så hyppig som den tilladte form med j, dvs. linjere.
→ linjere

[The Danish Language Board does not accept this spelling although it is more than 25 times as frequent as the allowed form with j, i.e. linjere.]

An analogous example can be provided in Sepedi. In the *New Sepedi Dictionary*, various variants have been included although they are considered incorrect by the *Northern Sotho Terminology and Orthography No. 4* (Wolff 1988) which contains the official spelling rules of Standard Sepedi compiled by the former Sepedi Language Board. One of these non-official variants is *thekesi* where the correct form should be *taxi*. (We thank Prof. M.J. Mojalefa for this information.) In order to help guide users who only know *thekesi*, to the prescribed form, an article such as the following would be useful:

thekesi

The Sepedi Language Board does not accept this variant although it is frequently used in existing texts and colloquial language.
→ taxi

(d) In the three previous cases, the treatment of the cross-reference lemma has been the same in terms of prescriptive and proscriptive dictionaries. In the third type, proscription prevails and recommends a given form against the prescriptive indications of the official language policy, frequently supported by a corpus analysis showing that the recommended form is much more commonly used than the officially prescribed form. The following example of this method is also taken from *Den Danske Netordbog*:

curriculum vitae *substansiv* <et, -et, -, -ene>

Dansk Sprognævn godtager ikke denne skrivemåde, men kun curriculum vitae. Denne ordbog anbefaler curriculum vitae, da det er den normale skrivemåde i danske tekster.

[The Danish Language Board does not accept this spelling, but only *curriculum vitae*. However, this dictionary recommends *curriculum vitae* as it is the normal spelling in Danish texts.]

(e) The fifth type is a variant of the former. In this case, a distinction is made between formal or scientifically correct language, on the one hand, and informal, colloquial language on the other. One such example can be found in the English-Spanish *Encyclopedic Dictionary of Gene Technology* (Bergenholtz et al. 1998) where the user finds the information that the expression *DNA digestion* is

not a term in the strict sense of the word, but laboratory jargon used instead of the much more complex expression *hydrolysis of peptide bonds in proteins or phosphodiester bonds in nucleic acids*:

DNA digestion *digestión f del AND*

The expression digestion means hydrolysis of peptide bonds in proteins or phosphodiester bonds in nucleic acids in laboratory jargon.

6.2 How to present the prescribed and proscribed forms

As to the words and variants prescribed or proscribed by the body responsible for the national, domain-specific or terminological language policy and selected as lemmas, there are at least five possible types of treatment:

- (a) Not giving a note with reference to the non-allowed or non-recommended variants,
- (b) Giving a note indicating the non-allowed or non-recommended variants,
- (c) Giving a note indicating that the non-allowed or non-recommended variants are frequently used,
- (d) Giving a note recommending the non-prescribed variants, and
- (e) Giving a note recommending the non-prescribed variants in colloquial language.

With reference to the discussion above, it is important to remember that no word selected as a lemma is in itself either prescribed or proscribed if this value is not directly stated in the article or indirectly mentioned, i.e. somewhere else in the dictionary. On this basis, the following comments can be made:

- (a) In most cases, the prescribed and proscribed variants are the only existing forms and it is therefore neither possible nor relevant to include a note about non-allowed or non-recommended variants. However, it is surprising to see that even if there are other non-prescribed or non-proscribed variants, some of which are also selected as cross-reference lemmas, these variants are not mentioned in the prescribed or proscribed articles of the majority of existing dictionaries. It is difficult to guess the reason for this decision, but it may be because the lexicographers think that the inclusion of 'non-desired' variants may confuse users who should only know about the prescribed or proscribed forms. However, the opposite argument could also be used. Users who have accessed the lemma in question through a non-allowed or non-recommended form may need confirmation that the found lemma actually corresponds to and has the same meaning as the one they first looked up, especially if the spelling differs much. This problem is especially big in electronic dictionaries where the users who enter a non-allowed or non-recommended variant in the search

machine are frequently guided directly to the prescribed and proscribed form. This is for instance the case with the official Danish orthographic dictionary *Retskrivningsordbogen* (Dansk Sprognævn 2006) compiled by the Danish Language Board in order to announce its decisions regarding correct spelling and inflection. If users enter the frequently used, but non-allowed variant *curriculum vitae*, they will be guided directly to the following article:

curriculum vitæ *sb.*, -et, curriculum vitæ,
bf. pl. -ene (fork.: CV el. cv).

Users who need information about inflection for text-production purposes may not even notice that the spelling is different from the one used to access the article, and in the subsequent text production they may repeat the spelling variant which the official Danish Language Board considers to be incorrect, but which they themselves consider the right one. It should therefore be recommended that electronic dictionaries include notes giving information that the prescribed or proscribed form is different from the one users entered into the search machine. The same holds true for direct orthographic mistakes which users may not be aware of.

(b) In the second type of treatment, the dictionary informs, in one way or another, that there are other variants than the recommended or allowed ones. For instance, the problem mentioned above could easily be solved in the following way:

curriculum vitæ *sb.*, -et, curriculum vitæ,
bf. pl. -ene (fork.: CV el. cv).
NB: You wrote *curriculum vitae* which is an incorrect spelling.

(c) In some cases, it could also be relevant to inform users that the non-allowed or non-recommended form is frequently used. An example which includes a note giving information about the frequent use of a non-allowed form is provided by the Danish Music Dictionary *Musikordbogen* (I. Bergenholtz 2006):

akkordeon
 et harmonika-instrument
 [...]

 Ved søgning på nettet ser det ud, som om betegnelsen med to c'er, som ikke anerkendes af Dansk Sprognævn, som regel bruges af konservatorierne i beskrivelsen af deres uddannelser, mens de to k'er ses i tekster fra musikskoler, hjemmesider mv.

 [Searching on the internet, it seems that the spelling with two c's, which is not allowed by the Danish Language Board, is usually used by the academies of music in the description of their training programmes, while the two k's are used in texts from music schools, home pages, etc.]

This article follows the official spelling rules laid down by the Language Board and, at the same time, indicates that a different spelling is frequently used by

academies of music. In this way it assists users who may be confused because they have met or only know the non-allowed spelling variant *accordeon* which is also frequently used by musicians.

(d) Sometimes the lexicographers decide, for some reason or other, to go against the official language policy and recommend a non-allowed variant. This is mostly the case when this variant is much more frequently used than the recommended one. An example of this type has already been provided with the Danish *curriculum vitae* in point (d) of paragraph 6.1.

(e) In other cases, it could be relevant to recommend the variant prescribed by the language board for use in official and formal language and another non-allowed variant to be used in informal and colloquial language. No example of this practice has been found, but it could be relevant in order to solve problems related to specific language communities, for example the problems regarding the big difference between the official and the informal language on the Faroe Islands.

6.3 How to present inflectional variants of the same word

In some languages, such as Danish and Afrikaans, the official language policy permits two or more inflectional patterns to the same word. This is for instance the case with the Danish verb *fnise* (English *giggle*) where the Danish Language Board permits three past tenses (*fnisede*, *fniste*, *fnes*) and two past participles (*fniset*, *fnist*) as can be seen in the following article from its official orthographic dictionary *Retskrivningsordbogen*:

fnise *vb.*, -ede *el.* -te *el.* fnes, fniset *el.* fnist.

This descriptive way of presenting the allowed inflectional forms may be acceptable in text reception when users just want to confirm that they have found the right lemma or in cognitive situations where users need general information about the word and its morphological pattern. However, in the case of text production, this way of presenting the data may confuse users because they have to choose between several inflectional variants which may not have the same frequency in daily language use. This problem can be solved by means of a proscriptive approach as the one followed by *Den Danske Netordbog* which is a dictionary conceived for text production:

fnise *verbum* <-r, -de, -t>
Dansk Sprognævn godtager også fniste/fnes, fnist.
[The Danish Language Board also allows fniste/fnes, fnist.]

With this proscriptive method, *Den Danske Netordbog* recommends the most frequently used inflectional forms (*fniser*, *fnisede*, *fniset*) — which the users who

are most often in a hurry can use immediately without reading the rest of the article — and at the same time indicates that the official language policy also allows other variants which users may prefer for stylistic or other reasons.

More than one inflectional possibility is also allowed for some Afrikaans words. In the *Afrikaanse Woordelys en Spelreëls*, the Language Commission allows two plural forms for many nouns ending in *-ing*, i.e. to add either an *-e* or an *-s* as can be seen in an example like *onderskeiding* (English *distinction*) and its plural forms *onderskeidinge/onderskeidings*. Both the *-e* and the *-s* plural forms have a high usage frequency and could be recommended in a proscriptive dictionary. However, there are other inflectional variants officially acknowledged by the Language Commission where there are definite differences in usage frequency. For nouns like *musik* (English *musician*) and *fisik* (English *physicist*) the plural forms *musikusse/musici* and *fisikusse/fisici* are acknowledged although the forms *musici* and *fisici* are much more frequent. This should be indicated in a dictionary with a proscriptive approach, e.g.:

fisikus (fisici) ...

Die Taalkommissie laat ook *fisikusse* as meervoudsvorm toe.

[The Language Commission also allows *fisikusse* as plural form.]

In terms of user-friendliness, such a proscriptive approach should always be recommended in dictionaries conceived to assist users in solving problems related to text production.

7. Conclusion

The formulation of language policy typically leads to the prescription of those forms recognised by the official language body as being correct or pure. Prescription states which forms should be used, excluding the use of all other forms. Description presents a variety of forms without differentiating between accepted and non-accepted, recommended and non-recommended, whereas prescription offers one or more recommended forms. Different types of prescription prevail and the application of certain types can also lead to an indication of the non-recommended forms.

This article emphasizes the importance of a distinction between different levels of language policy and the implications that decisions on these different levels have for lexicographers. Lexicographers need to deal with the prescription by official language bodies but can only give a thorough lexicographic account of language if they respond to the needs of their intended users and the envisaged functions of their dictionaries. In order to be able to do this, this article proposes that lexicographers should take cognisance of prescriptive decisions by language bodies but in their dictionaries they should venture beyond the prescriptive visions. By being proscriptive lexicographers make recommendations to their users and often these recommendations coincide

with the decisions prescribed by various official language bodies. Where the needs of dictionary users and the functions of a dictionary demand a deviation from the prescribed forms, lexicographers may recommend the non-prescribed forms that would best serve the purpose of their dictionary. Access to the recommended forms is important and where non-recommended forms, forms that are not allowed and even forms representing frequently made spelling mistakes are also included in a dictionary, it gives users additional access routes to the recommended forms.

Examples of the rigorous application of a prescriptive approach in dictionaries, a discussion of access problems and possibilities and examples of the application of proscription and the subsequent lexicographic presentation lead to the concluding recommendation of this article. Relevant data presentation, the best possible access to the data, an optimal retrieval of information and the satisfaction of the lexicographic functions of a dictionary are paramount to ensure successful dictionary use. To achieve this, lexicographers should not merely abide by the results of prescription. Dictionaries could do better by offering a wider selection of forms, complemented by well-motivated recommendations in order to guarantee the success of dictionaries as utility tools.

Literature

Dictionaries

- Bergenholtz, H. and U. Kaufmann (Eds.).** 1992. *Genteknologisk ordbog. Dansk-engelsk/engelsk-dansk molekylærbiologi og DNA-teknologi*. Copenhagen: Gads Forlag.
- Bergenholtz, H. and U. Kaufmann (Eds.).** 1998. *Encyclopedic Dictionary of Gene Technology. English-Spanish*. Toronto: Locus.
- Bergenholtz, H. et al. (Eds.).** 2008. *Den Danske Netordbog* [online]. Odense: Ordbogen.com. <http://www.ordbogen.com/opslag.php?&dict=ddno> [18 June 2008].
- Bergenholtz, I. (Ed.).** 2006. *Musikordbogen* [online]. Aarhus: Aarhus School of Business. <http://www.musikordbogen.dk> [18 June 2008].
- Bosman, D.B. et al. (Eds.).** 1984^s. *Tweetalige Woordboek/Bilingual Dictionary*. Cape Town: Tafelberg.
- Dansk Sprognævn.** 2006. *Retskrivningsordbogen på nettet* [online]. Copenhagen: Dansk Sprognævn. <http://www.dsn.dk> [18 June 2008].
- Eksteen, L.C.** 1997¹⁴. *Groot Woordboek Afrikaans-Engels/Engels-Afrikaans / Major Dictionary Afrikaans-English/English-Afrikaans*. Cape Town: Pharos.
- Hjorth, E. and K. Kristensen (Eds.).** 2003–2006. *Den Danske Ordbog. Volume 1–6*. Copenhagen: Gyldendal.
- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton/T. & T. Longman et al.
- Komensky, A.** 1631. *Ianua Linguarum Reserata*.
- Kriel, T.J.** 1983. *Pukuntšu Dictionary*. Pretoria: J.L. van Schaik.
- Ondal, F.F. and R.H. Gouws (Eds.).** 2000⁴/2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Johannesburg: Perskor/Cape Town: Pearson Education.

- Prinsloo, D.J. and B.P. Sathekge.** 1996. *New Sepedi Dictionary*. Pietermaritzburg: Shuter and Shooter.
- Schoonees, P.C. et al. (Eds.).** 1951–. *Woordeboek van die Afrikaanse Taal*. Pretoria: Government Printer/Stellenbosch: Bureau of the WAT.
- Silva, P. (Ed.).** 1996. *A Dictionary of South African English on Historical Principles*. Oxford: Oxford University Press.
- Taalkommissie van die S.A. Akademie.** 2001. *Afrikaanse Woordelys en Spelreëls*. Kaapstad: Pharos.
- Wolff, H.** 1988. *Northern Sotho Terminology and Orthography No. 4*. Pretoria: Government Printer.

Other literature

- Bergenholtz, H.** 2001. Proskription, oder: So kann man dem Wörterbuchbenutzer bei Textproduktionsschwierigkeiten am ehesten helfen. Lehr, A. et al. (Eds.). 2001. *Sprache im Alltag*: 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.** 2006. How to Do Policy with Dictionaries. *Lexikos* 13: 13-45.
- Bergenholtz, H., U. Kaufmann and S. Tarp.** 1994. Vore mænd i Havanna: Udarbejdelse af konception til en spansk-englisk genteknologisk ordbog. *Hermes, Journal of Linguistics* 13: 291-304.
- Bergenholtz, H. and S. Tarp.** 2007. Politik und Sprachpolitik in der Lexikographie. Gottlieb, H. and J.E. Mogensen (Eds.). 2007. *Dictionary Visions, Research and Practice*: 217-240. Amsterdam: John Benjamins.
- Gouws, R.H.** 1995. Dictionaries and the Dynamics of Language Change. Kachru, B. (Ed.). 1995. *Culture, Ideologies and the Dictionary*: 297-313. Tübingen: Max Niemeyer.
- Gouws, R.H.** 1999. Situating a Dictionary of South African English on Historical Principles within a More Comprehensive Lexicographic Process. *Lexikos* 9: 269-282.
- Gouws, R.H. and F.A. Ponielis.** 1992. The Development of Afrikaans and the Lexicographic Tradition. Zgusta, L. (Ed.). 1992. *History, Languages and Lexicographers*: 77-104. Tübingen: Max Niemeyer.
- Johnson, S.** 1747. *The Plan of a Dictionary of the English Language*. Facsimile Edition 1970. Menston: The Scholar Press.
- Mojela, V.M.** 2007. Standardization or Stigmatization? Challenges Confronting Lexicography and Terminography in Sesotho sa Leboa. *Lexikos* 18: 119-130.
- Nong, S., G.-M. de Schryver and D.J. Prinsloo.** 2002. Loan Words versus Indigenous Words in Northern Sotho: A Lexicographic Perspective. *Lexikos* 12: 1-20.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Max Niemeyer.
- Van der Merwe, M.F.** 2008. Wine and Words: A Trilingual Wine Dictionary for South Africa. *Lexikos* 18: 337-348.
- Wiegand, H.E.** 1986. Von der Normativität deskriptiver Wörterbücher. Zugleich ein Versuch zur Unterscheidung von Normen und Regeln. Henne, H. (Ed.). 1986. *Sprachnormen in der Diskussion. Beiträge vorgelegt van Sprachfreunden*: 72-101. Berlin/New York: De Gruyter.

Wörterbuchregister. Grundlagen einer Theorie der Register in modernen Printwörterbüchern

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg, Heidelberg, Bundesrepublik Deutschland* (herbert.ernst.wiegand@gs.uni-heidelberg.de)

Zusammenfassung: In diesem Beitrag werden die Begriffe bereitgestellt, die benötigt werden, um jedes bereits gegebene oder noch zu planende gedruckte Wörterbuchregister in einer einheitlichen theoretischen Perspektive analysieren und entwickeln zu können, und zwar hinsichtlich seiner Strukturen, seiner Funktionen und seiner Typzugehörigkeit. Dazu werden zunächst die akzessiven Registerinträge untersucht: Zahlreiche Typen von Registerinträgen werden unterschieden, wie z.B. nach der Anzahl und nach den Typen der Registerangaben, die reduzierten, die vollständigen, die einfachen, die erweiterten und die angereicherten Registerinträge sowie nach der mediostrukturellen Orientierung u. a. die einfach und die mehrfach außenorientierten Registerinträge. Bei den Registerintragsstrukturen werden hierarchische registerinterne Konstituentenstrukturen, hierarchische registerinterne Mikrostrukturen, registerinterne Adressierungsstrukturen sowie hierarchische registerinterne Angabenstrukturen dargestellt. Weiterhin wird eine Typologie von Registerzugriffsstrukturen präsentiert, in der u. a. die mediostrukturellen von den nichtmediostrukturellen Registerzugriffsstrukturen unterschieden und beide Typen weitgehend subtypologisiert werden. Auf der Basis der unterschiedenen Eigenschaften und Teilen von Registern wird schließlich eine Typologie von Registern entworfen. Abschließend werden die Registerfunktionen betrachtet und exemplarisch gezeigt, was an Registern kritikwürdig ist.

Stichwörter: AKZESSIVER REGISTEREINTRAG, EINFACHER REGISTEREINTRAG, ERWEITERTER REGISTEREINTRAG, EXTERNE DATENAKZESSIVITÄT, GESTAFFELTE REGISTERZUGRIFFSSTRUKTUR, HIERARCHISCHE REGISTERINTERNE MIKROSTRUKTUR, MEDIOSTRUKTURELLE REGISTERZUGRIFFSSTRUKTUR, MEDIOSTRUKTURELLES REGISTER, MONODIREKTIONALES ZUGRIFFSREGISTER, PERIPHERE ALPHABETISCHE REGISTERZUGRIFFSSTRUKTUR, POLYDIREKTIONALES ZUGRIFFSREGISTER, REGISTERFUNKTION, REGISTERINTERNE ADRESSIERUNGSSTRUKTUR, REGISTERINTERNE ANGABENSTRUKTUR, ZENTRALE ALPHABETISCHE REGISTERZUGRIFFSSTRUKTUR, REGISTEREINGANG, REGISTERVERWEISANGABE, REGISTERVERWEISEINTRAG, REGISTERZUGRIFFSSTRUKTUR, WÖRTERBUCHREGISTER, ZUGRIFFSREGISTER

Abstract: Dictionary Indexes. Foundations of a Theory of Indexes in Modern-day Printed Dictionaries. This article presents the concepts needed for the analysis and development of all existing and future dictionary indexes according to a uniform theoretical perspective regarding their structures, functions and typology. For this purpose, the accessible index entries are examined. Numerous types of index entries are distinguished, e.g. according to

the number and types of index items, the reduced, the complete, the single, the expanded and the enriched index entries, as well as according to the mediostructural orientation, among others the single and the multiple externally oriented index entries. For the index entry structures, hierarchical index internal constituent structures, hierarchical index internal microstructures, index internal addressing structures as well as hierarchical index internal item structures are introduced. Furthermore a typology of index access structures is presented in which, among others, mediostructural are distinguished from the non-mediostructural index access structures and both types are extensively subtypologised. On the basis of the different features and parts of indexes a typology of indexes is proposed. In conclusion, the functions of indexes are examined and, by using examples, the aspects of indexes worthy of criticism are shown.

Keywords: ACCESS INDEX, ACCESSIBLE INDEX ENTRY, CENTRAL ALPHABETICAL INDEX ACCESS STRUCTURE, DICTIONARY INDEX, EXPANDED INDEX ENTRY, EXTERNAL DATA ACCESSIBILITY, HIERARCHICAL INDEX INTERNAL MICROSTRUCTURE, INDEX ACCESS STRUCTURE, INDEX ENTRY, INDEX FUNCTION, INDEX INTERNAL ADDRESSING STRUCTURE, INDEX INTERNAL ITEM STRUCTURE, INDEX REFERENCE ENTRY, ITEM GIVING THE INDEX CROSS-REFERENCE, MEDIOSTRUCTURAL INDEX, MEDIOSTRUCTURAL INDEX ACCESS STRUCTURE, MONODIRECTIONAL ACCESS STRUCTURE, PERIPHERAL ALPHABETICAL INDEX ACCESS STRUCTURE, POLYDIRECTIONAL ACCESS STRUCTURE, SINGLE INDEX ENTRY, STAGGERED INDEX ACCESS STRUCTURE

1. Worum es geht und warum nicht

In diesem Beitrag zu einem spezifischen Ausschnitt aus der Theorie der Wörterbuchform von Printwörterbüchern geht es darum, die lexikographietheoretischen Begrifflichkeiten und zugehörigen deutschen Termini sowie ihre systematischen terminologiesemantischen Beziehungen geordnet bereitzustellen, die benötigt werden, um jedes bereits gegebene oder noch zu planende gedruckte Wörterbuchregister in einer einheitlichen theoretischen Perspektive hinsichtlich seiner registerinternen Strukturen, hinsichtlich seiner Registerfunktionen und seiner typspezifischen Art als Benutzerschnittstelle zu funktionieren sowie hinsichtlich seiner Typzugehörigkeit beschreiben und planen zu können. Anders ausgedrückt heißt das: Es wird angestrebt, die Grundlagen für eine Theorie der Register in gedruckten Wörterbüchern zu entwickeln.

Ein solches Vorhaben ist gerade in der gegenwärtigen Entwicklung der Lexikographie insofern aktuell, als alle Formen von digitalen Wörterbüchern variable und vielseitige externe Zugriffsmöglichkeiten aufweisen, und die Printwörterbücher besonders hinsichtlich der externen Datenakzessivität (i.S.v. Wiegand 2005 u. 2008) stärkeren Beschränkungen unterworfen sind: Beschränkungen aufgrund des Datenträgers Papier, solchen aufgrund der zu wählenden Datenfixierungssysteme und nicht zuletzt solchen aufgrund eingespielter starrer lexikographischer Traditionen (vgl. u.a. Wiegand 1995: 464f). Bereits in Wiegand (1983: 442 u. 464) habe ich das Vorgehen der Lexikographen, die große Mengen von mühsam erhobenen und wertvollen Sprachdaten (relativ

stupide) so organisieren, dass sie nur über gerade eine alphabetische makrostrukturelle Zugriffsstruktur erreichbar sind — wie z.B. im Duden-GW — ironisch die „Methode der lexikographischen Datentarnung“ genannt. Damit sollte allerdings keineswegs geleugnet werden, dass Printwörterbücher nur als statische Informationssysteme hergestellt werden können. *Statisch* meint hier vor allem, dass die Zugriffsstrukturen aller Typen (vgl. zu diesen Wiegand 2008) textuell stets fest implementiert sind, so dass zu jedem Zugriffsstrukturtyp meistens nur ein Zugriffspfadtyp (i.S.v. Wiegand 2007) und nur selten mehrere gehören. *Statisch* heißt aber nicht, dass Printwörterbücher nur gerade eine äußere Zugriffsstruktur aufweisen müssen, so dass sie zum Typ des einfach direkt monoakzessiven Wörterbuchs gehören. Vielmehr kann die externe Datenakzessivität durch Register, die bestimmten Wörterbuchfunktionen zugeordnet sind (bzw. sein sollten), ganz erheblich verbessert werden. Daher ist es relevant, auf systematische Weise zu wissen und theoretisch zu verstehen, welche Möglichkeiten für die Optimierung der externen Datenakzessivität von Printwörterbüchern mit welchen Typen von Registern prinzipiell gegeben sind.

Ausdrücklich sei nun auch gesagt, worum es in diesem Beitrag nicht geht: Nicht behandelt wird der weitgefächerte Fragenkomplex, wie Register von Hand oder automatisch hergestellt werden: Die so genannte Indexierung ist ein anderes dokumentationswissenschaftliches Thema. Weiterhin verfolgt dieser Beitrag keine lexikographiehistorischen Ziele, so dass nicht darauf eingegangen wird, wie Frühformen von Wörterbuchregistern gestaltet waren und welche Funktionen sie hatten (vgl. dazu: Zedelmaier 2004; Blair 2000; Rouse und Rouse 1982).

Schließlich sei zum Abschluss dieser Vormerkung noch *expressis verbis* darauf hingewiesen, dass das hier skizzierte Theoriemodul zu einer Allgemeinen Theorie der Lexikographie gehört und mit anderen Modulen interagiert (vgl. Wiegand 1998: 1-10). Es ist daher nicht möglich, alle verwendeten Termini aus anderen Theoriemodulen erneut ausführlich zu erklären. Meistens können nur kurze erklärende Hinweise gegeben werden, und öfters kann ich nur auf andere Publikationen verweisen.

2. Bauteile, Strukturen und Typen von Wörterbuchregistern

Jeder, der Wörterbücher benutzt hat, weiß auch mehr oder weniger genau, was ein Register in einem gedruckten Wörterbuch ist. Hat man jedoch ein „Register“ zu 12 Leitzordnern mit Kopien der ersten drei Seiten von „Verzeichnissen“ aus über 500 Wörterbüchern (als relativ schmale empirische Basis für diese Untersuchung), dann bemerkt man allmählich, dass das praxisbasierte am prototypischen Fall orientierte, alltägliche Wissen über Register nicht einmal ausreicht, um in jedem Fall mit guten Gründen entscheiden zu können, ob ein vorgelegtes wörterbuchinternes „Verzeichnis“ als Wörterbuchregister gelten soll oder nicht. Das fängt damit an, dass Wörterbuchregister unter recht unterschiedlichen Bezeichnungen firmieren, wie z.B. *Register*, *Stichwortregister*, *Ele-*

mentenregister, Verzeichnis der ... (z.B. „Alphabetisches Verzeichnis der lateinischen Ursprungswörter“ in Kytzler und Redemund 2007 oder „Verzeichnis nicht an alphabetischer Stelle behandelter Wörter“ in Pfeifer 1993) weiterhin Bezeichnungen wie *Index, Index verborum, Index rerum, Liste der ..., Rangliste der ..., Zugriffsregister, vollständiges Zugriffsregister, Umkehrwörterbuch* u.a. sowie solche vom Typ „LEMMATA SACHLICH GEORDNET“ (z.B. „Entlehnungen nach Fachgebieten geordnet“ in Pfeffer 1987: 21-40), und es hört damit auf, dass zwar das prototypische Wörterbuchregister im Nachspann eines Wörterbuchs zu finden ist, dass es aber auch „Verzeichnisse“ im Vorspann gibt, wie z.B. das „Sachregister — Wegweiser zum Wortschatz“ in Paul (2002) oder das „Register zur lexikographisch-historischen Einführung“ in Dornseiff (2004) sowie weiterhin lexikographische Register in Außentexten (vgl. Wiegand 2008), die nicht zum Buchblock eines Printwörterbuchs gehören. Aus diesen Gründen hat es daher wenig Sinn, mit einer Definition von *Wörterbuchregister* zu beginnen. Wir beginnen vielmehr mit exemplarischen Analysen prototypischer Wörterbuchregister (im Folgenden kurz: Register) mit dem Ziel der fortschreitenden und sich schrittweise differenzierenden Begriffsbildung.

2.1 Akzessive Registereinträge: Bauteile, Typen und eintragsinterne Strukturen

Akzessive Einträge (die im Englischen (*basic reference unit*) heißen (vgl. z.B.: Hartmann und James 1998 u. Wiegand 2003: 173ff)) sind die textuellen Basiseinheiten jedes gedruckten Nachschlagewerkes. Sie treten in nichtlexikographischen Nachschlagewerken wie Warenkataloge, Telephonbüchern, Bibliographien und von ihren Besitzern geführten Terminkalendern ebenso auf wie in allen Printwörterbüchern aller denkbaren Typen. Man unterscheidet folgende Typen von wörterbuchinternen akzessiven Einträgen: Akzessive Binnentexteinträge, akzessive Einschubeinträge, akzessive Umtexteinträge, akzessive Registereinträge und als die wichtigsten die akzessiven Wörterverzeichnis-einträge, die meistens *Wörterbuchartikel* genannt werden. Zu den lexikographischen akzessiven Einträgen gehören weiterhin die wörterbuchexternen akzessiven Außentexteinträge sowie die akzessiven Listeneinträge (vgl. Wiegand 2008). Für jeden lexikographischen akzessiven Eintrag gilt das Folgende: Er weist mindestens ein äußeres Zugriffstextelement auf, das meistens initial in der ersten Eintragszeile steht (aber nicht immer, vgl. (7) in Abb. 1) und ein Element einer äußeren Zugriffsstruktur ist. Bei Wörterbuchartikeln in alphabetischen Wörterbüchern heißt das äußere Zugriffstextelement *Lemma*. Lemmata sind Leitelementträger. Bei nichtalphabetischen Wörterbüchern heißt das äußere Zugriffstextelement *Wörterbuchartikeltitel* (kurz: *Artikeltitel*). Artikeltitel sind keine Leitelementträger, sondern Verweisadressenträger; auf sie kann nur registervermittelt zugegriffen werden. Bei allen anderen akzessiven Einträgen liegen als äußere Zugriffstextelemente Eintragseingänge vor, die *Binnentexteingang, Einschubeingang, Umtexteingang, Registeringang, Außentexteingang* und

Listeneingang heißen. Eintragseingänge können sprachlich, numerisch oder alphabetisch sein. Alle sprachlichen Eintragseingänge sind zugleich Zeichengestaltungangaben. Es gilt weiterhin, dass ein lexikographischer akzessiver Eintrag eine Konstituente eines lexikographischen Teiltextes mit äußerer Teiltextzugriffsstruktur ist, der akzessive lexikographische Daten enthält, und zwar primär akzessive und sekundär akzessive.

2.1.1 Bauteile und Typen von Registereinträgen

Im Folgenden betrachten wir anhand der Registerauszüge in Abb. 1 und Abb. 2 die wichtigsten Bauteile von Registereinträgen und unterscheiden verschiedene Typen von akzessiven Registereinträgen (re).

- | | | |
|--|--|--|
| <p>(1)</p> <ul style="list-style-type: none"> bischen s. beruhigen biss(e)l s. bißchen Bißgurre, -gurn s. Steit Bitz s. bißchen, Kerngehäuse Bitzel s. bißchen bitzeln s. prickeln Bitzlerl s. bißchen Bitzler s. prickeln Bitzgi s. Kerngehäuse Bitzli s. bißchen Blache s. Plane Blag(e) s. Kind blaken s. Ruß bläken s. brüllen blarren s. weinen | <p>(2)</p> <ul style="list-style-type: none"> Blase s. Wasserkessel Blatt s. Fuß Blatter s. Anschwellung Blattern s. Pocken Blätzer s. Narbe Blaubeere s. Heidelbeere Blaukabis } s. Kohl Blaukappes } s. Kohl Blaukraut } s. Kohl Blech s. reinigen Blechner } s. Klempner Blechschräger } s. Klempner Blechschrämied } s. Klempner Bletz(e) s. Narbe Bloch s. Block | <p>(3)</p> <ul style="list-style-type: none"> para-, Para- -paradies -pflichtig -phil -phob Pilot- Plüsch- Polit- -politisch poly-, Poly- post- prä- pro-, Pro- Problem- |
| <p>(4)</p> <p>Paracelsus ↑ <i>Begriffs- und Bedeutungsprägung</i></p> <p>Partikel ↑ <i>Gesprächswort</i> ↑ <i>Gradpartikel</i></p> <p>Pejoration ↑ <i>Bedeutungsverschlechterung</i></p> <p>Pejorativbildung auch ↑ <i>Schimpfwort</i> ↑ <i>Spottnamen</i> Dichterling · Empfinderei · Emporkömmling · Fatzke · Finsterling · Frömmling · Geizhals · Günstling · hündisch · kindisch · Kriegsgewinnler · Liebelei · Lüstling · Meckerfritze · Ossi · Piefke · Prahlhans · Raffke · Raufbold · Schreihals · Steppke · Trunkenbold · Weibsbild, -person, -stück · Weichling · Wessi · Witzbold · Wühler</p> <p>Persisch Babuschen · Basar · Bombast · Borax · Derwisch · Diwan · Karawane · Khaki · Kiosk · Limonade · Paradies · Rochade · Saffian · Schach · Schalscharlach · Spinat · Taft · Tasse · Tulpe · Turban · Zinnober</p> | <p>(5)</p> <p>Regiepult 14.10 Film und Kino</p> <p>regierbar 9.52 Leicht</p> <p>regieren 15.74 Führung; 18.12 Herrschen</p> <p>regierend 18.12 Herrschen</p> <p>Regierung 15.71 Schicht; 15.74 Führung; 18.4 Politik; 18.9 Regierung; 18.10 Verwaltungsbehörde; 18.12 Herrschen</p> <p>Regierungs- 18.10 Verwaltungsbehörde</p> | |

(6)			
backen 111			
bauen 130			
beachten 102			
beantworten 106			
bearbeiten 109			
beauftragen 285			
bedanken, sich 171			
bedrängen 214			
bedrohen 205			
beeilen, sich 231			
beenden 389			
beerben 281			
	(7)		
	anziehen (Kleidung)	H	anlegen
	* Anzihsachen	H	Gewand, Kluft, Montur
	* Apfelsine	H	Orange
	Appetit	A H	[i] kurz
	* Ar, der	H	Ar, das
	Arbeitspause	H	Brotzeit
	ärgern (es ärgert mich)	H;	das stinkt mir; mir stinkt er
	Arschbacke, die	H	Arschbacken, der
	* artig	H	brav
	As(s) (Spielkarte), das	H	die Ass, Sau
	* Aschkasten	H	Aschenschuber
	* Aspik	A H	[i] kurz: Sulz
	ö Ass	H	Eiß
	Ästchen		Äst(er)l, Zweigerl
	Atem		Schnaufer(er)
	atzen	H;	füttern; futtern, ämmeln
	** au Backe!		auweh zwick!

Abb. 1: Registerausschnitte (1) und (2) aus dem nachspanninternen alphabetischen Register in Seibicke (1989) mit dem Registertitel „Register der in den Wörterbuchartikeln erwähnten gemeinsprachlichen und landschaftlichen Ausdrücke“; (1) = re₁–re₁₅; (2) = re₁₆–re₂₆. Registerausschnitt (3) aus dem vorspanninternen reduzierten alphabetischen Register in Duden-10 mit dem Registertitel „Liste der Wortbildungselemente, die in der alphabetischen Stichwortliste erscheinen“; (3) = re₂₇–re₄₀. Registerausschnitt (4) mit dem Registertitel „Sachregister — Wegweiser zum Wortschatz“ aus dem vorspanninternen alphabetischen Register in Paul (2002); (4) = re₄₁–re₄₅. Registerausschnitt (5) aus dem nachspanninternen alphabetischen Register mit dem Registertitel „Vollständiges alphabetisches Zugriffsregister“ in Dornseiff (2004); (5) = re₄₆–re₅₂. Registerausschnitt (6) aus dem nachspanninternen alphabetischen Register mit dem Registertitel „Alphabetisches Sachregister“ in Helbig und Schenkel (1980); (6) = re₅₃–re₆₄. Registerausschnitt (7) aus dem nachspanninternen alphabetischen Register mit dem Registertitel „Das Umkehrwörterbuch. Einheitsdeutsch — Bairisches Deutsch“ in Zehetner (1997); re₆₅–re₈₁.

Der erste Registereintrag re₁ in (1) „bischen s. beruhigen“ ist ein vollständiger Registereintrag. Er besteht aus dem (sprachlichen) Registerzugang „bischen“, mit dem das Verb *bischen* genannt wird, und der unmittelbar folgenden nicht-elementaren Registerangabe „s. beruhigen“. Diese ist eine vollständige Verweisangabe (*sensu* Wiegand 2002: 186ff), und zwar eine vollständige Registerverweisangabe, weswegen re₁ und alle anderen Registereinträge in (1) zum Typ des vollständigen Registereintrags gehören. Die registerinterne vollständige Verweisangabe „s. beruhigen“ ist funktional-positional in zwei unmittelbare Teilangaben segmentierbar: Die erste Teilangabe ist die verdichtete Verweisbeziehungsangabe „s.“, mit der eine Abkürzung von *siehe* genannt wird; die zweite Teilangabe ist die Verweisadressenangabe „beruhigen“, mit der die Verweisaußenadresse |beruhigen| genannt wird, deren Verweisadressenträger das Lemma „beruhigen“ in wa₁ in Abb. 2 ist.

wa₁: **beruhigen**, *beschwichtigen*: niederdt. umgangssprachlich (*jemanden begöschten*; mittel- und süddt. *bischen* = ‚einen weinenden Säugling beruhigen, auf den Armen wiegen‘.

wa₂: **interimsabschied, interimsaufwartung, interimsbesitzer, interimscondition, interimseinteilung, interimsunterhaltung, interimsvermittlung, interimsverordnung, interimsweise, interimswesen**, s. *interim* 1.

(8)
Entlehnung ↑ *Bedeutungsentlehnung* ↑ *Fremdwortverdeutschung* ↑ *Lehnbildung* ↑ *Lehnschöpfung* ↑ *Lehnübersetzung* ↑ *Lehnübertragung* ↑ *Lehnwort* ↑ *Mehrfachentlehnung* ↑ *Rückentlehnung* ↑ *Scheinentlehnung* ↑ *Wortexport*
Entzückungswort *auch* ↑ *Jugend-sprache* ↑ *Schülersprache* ↑ *Studentensprache*
Epochenbezeichnung
Ersatzwort ↑ *Fremdwortverdeutschung* ↑ *Lehnschöpfung* ↑ *Lehnübersetzung* ↑ *Lehnübertragung*

(9)
Birkler, Hubertus → Ettle, Josef
Birkmoser, Hanni → Falk, Hermann
Birkner, Friede → Stein, Elfriede
Birkner, Polten → Brosowski, Paul
Birmann, Martin → Grieder, Martin
Birnitz, A. von → Czibulka, Alfons von
Birnitz, Linda → Pietzsch, Lina
Biron, Hans → Brüstle, Hans
Biron, Ludwig → Siegert, Georg

(10)
Leibniz, *Gottfried Wilhelm* 9*
Leitelement 32*, 64*
Leitelementträger 64*
Lektüreziel, offenes 42*
Lemmazeichen 64*
Lemmazeichengestaltangabe 85*
Lewy, *E.* 23*
lexikographische Information
s. Information, lexikographische
– Makrostruktur
s. Makrostruktur, lexikographische
– Wörterbuchpflege
s. Wörterbuchpflege, lexikographische

Abb. 2: Wörterbuchartikel (= wa) wa₁ aus Seibicke (1989); verdichteter Verweisartikel wa₂ aus FWB 8/1; Registerausschnitt (8) aus Paul (2002); (8) = re₈₂–re₈₅. Registerausschnitt (9) aus Eymmer (1997); (9) = re₈₆–re₉₄. Registerausschnitt (10) aus dem Register mit dem Registertitel „Register zur lexikographisch-historischen Einführung“ in Dornseiff (2004); (10) = re₉₅–re₁₀₄.

Jeder sprachliche Registereingang ist ein Leitelementträger; d.h. am Beispiel von re₁: Das Verb *bischen* liefert einem Benutzer das Leitelement |bischen|, mit

dessen Hilfe er eine Benutzungshandlung vom Typ der externen Registerzugriffshandlung (*sensu* Wiegand 1998: 472f) ausführen kann, um in der Registerzugriffsstruktur den Registerzugang „bischen“ zu finden, der das gleiche Leit-element wie das Verb *bischen* aufweist. In vollständigen Registerinträgen kann die Verweisbeziehungsangabe auch durch Verweisungspfeile realisiert sein, wie z.B. in (9) in Abb. 2.

Der erste Registerbeitrag re_{53} in (6) in Abb. 1 „backen 111“ ist ein reduzierter Registerbeitrag: Die registerinterne numerische Verweisangabe „111“, mit der als Außenadresse eine Seitenzahl als Element der äußeren Seitenzahlzugriffsstruktur genannt wird, gehört zu den elementaren reduzierten Verweisangaben (*sensu* Wiegand 2002: 186ff), die als Teilangaben keine Verweisbeziehungsangaben aufweisen. Reduzierte Verweisangaben (die sich auch in Wörterbuchartikeln finden) bestehen nur aus einer Verweisadressenangabe. Alle reduzierten Registerbeiträge, mit deren Registerangaben Seitenzahlen oder Spaltenzahlen genannt werden, funktionieren wie eine reduzierte Doppeladressenangabe. Mit der Außenadressenangabe „111“ wird die numerische Hauptadresse genannt, die in der Seitenzahlzugriffsstruktur liegt; mit dem Registerzugang wird die Unteradresse |backen| genannt, deren Verweisadressenträger der Artikeltitle **backen** ist, auf den nur registervermittelt zugegriffen werden kann (da Helbig und Schenkel 1980 ein nichtalphabetisches Wörterbuch ist). Ein Benutzer-in-actu, der anhand des reduzierten Registerbeitrages „backen 111“ einen Registerverweis (i.S.v. Wiegand 2002: 211ff) erschließt und diesem durch die Ausführung einer Verweisbefolgungshandlung folgt, muss zuerst auf die Hauptadresse, nämlich auf die Seitenzahl 111 zugreifen und kann erst danach die Unteradresse und damit **backen** auf der Seite 111 erreichen.

Die Registerbeiträge in (3) in Abb. 1 gehören alle zum Typ des vollständig reduzierten Registerbeitrages. Registerbeiträge dieses Typs bestehen nur aus mindestens einem Registerzugang. Wie sie funktionieren, wird in Abschnitt 2.2 erklärt.

Nicht mit allen numerischen registerinternen Außenadressenangaben werden Seiten- oder Spaltenzahlen genannt. In dem reduzierten einfachen Registerbeitrag „csatol 54“ aus (23) in Abb. 15 wird mit der numerischen Außenadressenangabe „54“ eine Lemmanummer genannt, da Ildikó (2007) eine alphanumerische makrostrukturelle äußere Doppelzugriffsstruktur aufweist, so dass die Lemmata mit Lemmanummernangaben durchnummeriert sind (vgl. Wiegand 2008). In dem reduzierten einfachen Registerbeitrag „brevitas 0436“ aus (24) in Abb. 15 wird mit der numerischen Außenadressenangabe „0436“, die in der Zugriffsspalte des tabellarischen Wörterverzeichnis von Kytzler und Redemund (2007) stehende Zugriffsnummer genannt, die die Artikelblöcke zählt (vgl. Wiegand 2000a).

Im Folgenden betrachten wir re_{41} , den ersten Registerbeitrag in (4), der folgende Form aufweist: „**Paracelsus** ↑ *Begriffs- und Bedeutungsprägung*“. In der Vorbemerkung des Registers mit dem Registertitel „Sachregister — Wegweiser

zum Wortschatz“ heißt es in Paul (2002): „„↑“ verweist auf Stichwörter im Sachregister“. Aufgrund dieser Erklärung gehört die durch einen Hochpfeil realisierte Verweisbeziehungsangabe für den kundigen Benutzer zum Typ der adressenidentifizierenden Verweisbeziehungsangabe: Die Verweisadresse wird als eine Adresse kenntlich gemacht, deren Verweisadressenträger ein Registereingang ist, nämlich der Registereingang „**Begriffs- und Bedeutungsprägung**“. Der Registereintrag re₄₁ entspricht einem reinen Verweisartikel. Er gehört daher zum Typ des Registerverweiseintrages, zu dem auch re₄₃ aus (4) in Abb. 1 gehört. Sowohl re₄₁ als auch re₄₃ weisen gerade eine vollständige Registerverweisangabe auf. Der Registereintrag re₄₂ „**Partikel** ↑ *Gesprächswort* ↑ *Gradpartikel*“ weist zwei Registerverweisangaben auf. Manche Registerverweiseinträge in Paul (2002) weisen mehr als zehn Registerverweisangaben auf. Es ist daher zweckmäßig, zwischen einfachen und erweiterten Registerverweiseinträgen zu unterscheiden. Einfache Registerverweiseinträge weisen gerade eine Registerverweisangabe auf (wie re₄₁ und re₄₃), erweiterte (wie re₄₂) dagegen mindestens zwei.

In dem Registerausschnitt (5) treten nur reduzierte Registereinträge auf. Mit den registerinternen reduzierten Verweisangaben, z.B. mit „9.52 Leicht“ aus re₄₇ „**regierbar** 9.52 Leicht“, wird gerade eine zweiteilige Verweisadresse genannt. Der Verweisadressenträger ist der Artikeltitel, bestehend aus der Sachgruppennummer „9.52“ und dem unmittelbar folgenden Artikeltitel „Leicht“, so dass keine Doppeladresse vorliegt, da Haupt- und Unteradressen im Verweiszielbereich niemals unmittelbar auf einander folgen. In re₄₈, dem Registereintrag mit dem Registereingang „**regieren**“, werden zwei Verweisadressen genannt, und in re₅₀, dem Registereintrag mit dem Registereingang „**Regierung**“, werden sechs Verweisadressen genannt. Es ist daher zweckmäßig, zwischen einfachen und erweiterten reduzierten Registereinträgen zu unterscheiden: Reduzierte Registereinträge, in denen nur eine Verweisadresse genannt wird, gehören zum Typ des einfachen reduzierten Registereintrags. Reduzierte Registereinträge, in denen mindestens zwei Verweisadressen genannt werden gehören dagegen zum Typ des erweiterten reduzierten Registereintrags, wie z.B. re₄₅ in (4) mit dem Registereingang „**Persisch**“. Eine analoge Unterscheidung ist auch bei den vollständigen Registereinträgen zweckmäßig. Beispielsweise gehört re₄ in (1) in Abb. 1, nämlich „Bitz s. bisschen, Kerngehäuse“ zum Typ des erweiterten vollständigen Registereintrags, während die anderen Registereinträge in (1) zum Typ des einfachen vollständigen Registereintrags gehören. Anhand von re₄ kann man übrigens erkennen, dass man den Unterschied zwischen einfachen und erweiterten Registereinträgen nicht ohne weiteres an der Zahl der Verweisadressenangaben festmachen darf, denn „s. bißchen, Kerngehäuse“ hat als Teilangabe eine (nichtelementare) homosegmentäre Verweisadressenangabe, die aus zwei Adressenangaben besteht, so dass mit einer nichtelementaren Adressenangabe zwei Verweisadressen genannt werden. Der Unterschied zwischen einfachen und erweiterten Registereinträgen muss mithin entweder (wie geschehen) an der Zahl der genannten Ver-

weisadressen oder an der Anzahl der elementaren Verweisadressenangaben festgemacht werden.

Bei allen drei charakterisierten Typen von Registereinträgen, dem Typ des erweiterten vollständigen, und dem des erweiterten reduzierten Registereintrags sowie dem Typ des erweiterten Registerverweiseintrags lassen sich die einfach von den mehrfach erweiterten Einträge unterscheiden. Eine Mehrfach-erweiterung liegt dann vor, wenn in einem Registereintrag mindestens drei Verweisadressen genannt werden. Mehrfach erweiterte reduzierte Registereinträge sind re_{45} aus (4) mit dem Registereingang „**Persisch**“ und re_{50} aus (5) mit dem Registereingang „**Regierung**“. Mehrfach erweiterte vollständige Registerverweiseinträge sind z.B. re_{82} aus (8) in Abb. 2 mit dem Registereingang „**Entlehnung**“ und re_{85} mit dem Registereingang „**Ersatzwort**“.

In (7) in Abb. 1, dem Registerausschnitt aus Zehetner (1997), treten links-erweiterte Registereinträge auf; die Linkserweiterung ist entweder durch einen Asterisk oder durch zwei realisiert oder durch „ö“. In allen Fällen handelt es sich um eine topikalisierte Registerangabe, die an den Registereingang adjazent rechtsadressiert ist. Mit „ö“ wird z.B. eine verdichtete Austriazismusidentifizierungsangabe realisiert. Die Registereingänge in (7), die topikalisierte Registerangaben aufweisen, gehören zum Typ des reduzierten Registereintrags mit topikalisierte Registerangabe.

Weiterhin ist unbedingt das Folgende zu beachten: Die Erweiterung und die Reduzierung betreffen nur die mediostrukturellen Eigenschaften und Teile von Registereinträgen und damit auch die Fundortdaten. Es gibt jedoch auch eine Anreicherung von Registereinträgen; diese bezieht sich nicht auf die Fundortdaten, sondern besteht darin, dass Registerangaben auftreten, die keine vollständigen oder reduzierten Verweisangaben sind. In (7) sind alle topikalisierten Registerangaben sowie die gestalkonstanten Angaben „A“, „B“ und „H“ keine Verweisangaben, sondern verdichtete Angaben: Mit „A“ wird z.B. angegeben, dass ein Ausspracheunterschied vorliegt und mit „B“, dass ein Unterschied in der Betonung gegeben ist. Registereinträge, die Angaben aufweisen, die keine Fundortangaben sind, heißen *angereicherte Registereinträge*. Alle Registereinträge in (7) in Abb. 1, die eine der drei topikalisierten Registerangaben aufweisen und/oder eine der beiden Angaben „A“, „B“ und/oder beide Angaben „A“ und „B“ sind angereicherte Registereinträge. Registereinträge mit gerade einer nicht verweisvermittelnden Angabe, wie z.B. „** au Backe! auweh zwick!“ (= re_{81}), sind einfach angereicherte Registereinträge. Registereinträge mit mehreren nichtverweisvermittelnden Angaben, wie z.B. „ö Ass H Eiß“ (= re_{77}), gehören zum Typ des mehrfach angereicherten Registereintrags. Da eine der nichtverweisvermittelnden Angaben, nämlich „ö“ eine topikalisierte Registerangabe ist, gehört re_{77} zum Typ des mehrfach angereicherten Registerintrags mit topikalisierte Registerangabe. Auch erweiterte Registereinträge können zusätzlich angereichert sein, so dass erweiterte und angereicherte Registereinträge vorliegen. — Im Folgenden fassen wir einen Teil der bisherigen Betrachtungen zu den Bauteilen und Eigenschaften von Registereinträgen zu einem Typologieausschnitt zusammen.

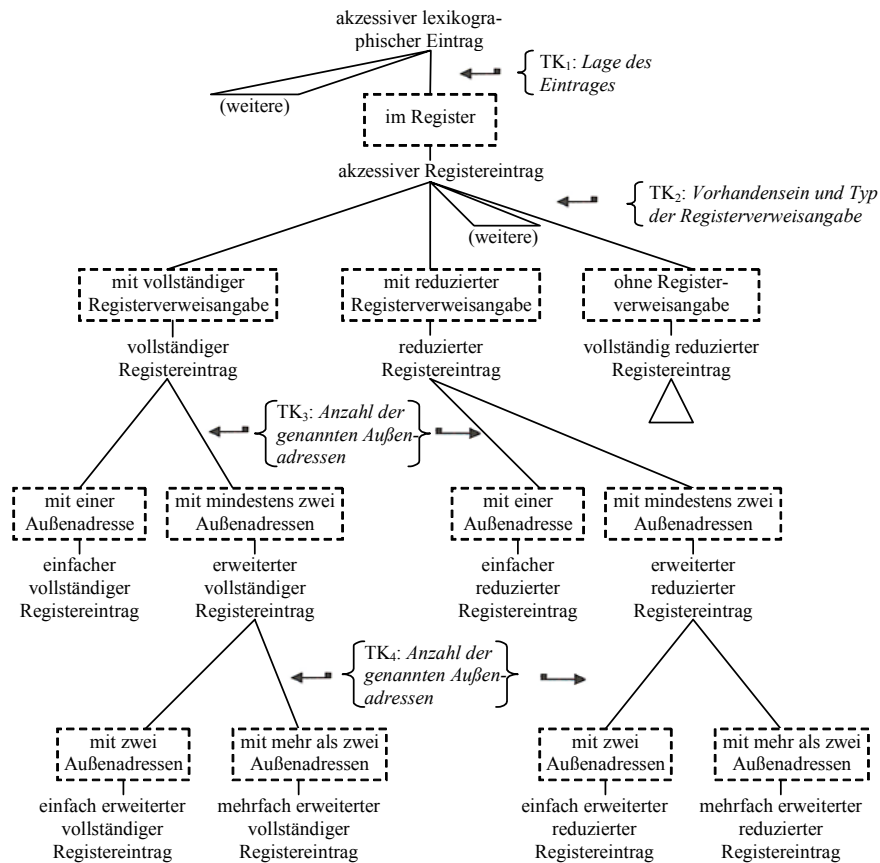


Abb. 3: Erweitert kommentierter Typologiegraph zum ersten Typologieausschnitt aus einer Typologie von Registerangaben in Printwörterbüchern nach Art und Anzahl der Registerverweisangaben. *Abkürzungen:* TK = Typologiekriterium; „ \rightarrow “ und „ \leftarrow “ bedeuten soviel wie die Anwendung des TK führt zu der Unterteilung

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3 gehören, seien die beiden Folgenden genannt:

- (D-1: einfacher vollständiger Registereintrag)
Ein einfacher vollständiger Registereintrag ist ein Registereintrag, auf dessen Registerzugang genau eine vollständige Registerverweisangabe folgt, deren Außenadressenangabe elementar ist.
- (D-2: erweiterter reduzierter Registereintrag)
Ein erweiterter reduzierter Registereintrag ist ein Registereintrag, auf dessen Registerzugang mindestens zwei reduzierte Registerverweisangaben folgen.

Im Folgenden betrachten wir die mediostrukturelle Orientierung der Registerverweisangaben und der zugehörigen Registereinträge. Jedes verweisvermittelnde Textsegment weist eine mediostrukturelle Orientierung auf (vgl. Wiegand 2002: 182ff). Diese wird durch das adressenvermittelnde Textsegment festgelegt, bei Verweisangaben mithin durch die Verweisadressenangabe. Es gibt mehrere Untertypen des Typs der Verweisadressenangabe; der wichtigste Untertyp ist der Typ der Außenadressenangabe. Die vollständige Registerverweisangabe „s. beruhigen“ ist lemmatisch orientiert, weil der Verweisadressenträger ein Lemma ist. Statt *lemmatisch orientiert* sagt man auch *artikelorientiert*. Ist der Verweisadressenträger ein Artikeltitel, sagt man immer *artikelorientiert*. Die mediostrukturelle Orientierung eines verweisvermittelnden Textsegmentes vererbt sich auf den akzessiven Eintrag, zu dem das verweisvermittelnde Textsegment gehört. Wird mit dem verweisvermittelnden Textsegment eine Verweisaußenadresse (kurz: Außenadresse) genannt, also eine, die nicht im gleichen akzessiven Eintrag zu finden ist, in dem das verweisvermittelnde Textsegment steht, dann weist das verweisvermittelnde Textsegment eine mediostrukturelle Außenorientierung auf. Wird dagegen mit dem verweisvermittelnden Textsegment eine Verweispinnenadresse (kurz: Binnenadresse) genannt, also eine, die im gleichen akzessiven Eintrag zu finden ist wie das verweisvermittelnde Textsegment, dann weist das verweisvermittelnde Textsegment eine mediostrukturelle Binnenorientierung auf. Binnenorientierte Registerverweisangaben sind zwar möglich, in meinem empirischen Material jedoch bisher nicht aufgetreten; d.h. registereintragsinterne Verweisung ist derzeit nicht belegbar; registerinterne Verweisung tritt dagegen öfters auf.

Die registerinterne vollständige Verweisangabe „s. beruhigen“ ist lemmatisch und damit außenorientiert. Diese mediostrukturelle Orientierung vererbt sich auf den Registereintrag. Demgemäß ist re_1 = „bischen s. beruhigen“ ein lemmatisch und damit außenorientierter Registereintrag. Da in re_1 nur eine Außenadresse genannt wird, ist re_1 ein einfach außenorientierter Registereintrag. Ein einfach außenorientierter Registereintrag ist immer monodirektional, so dass gilt: Alle Benutzer, die anhand von re_1 einen Verweis erschließen und eine Verweispflichthandlung erfolgreich ausführen, etablieren einen Zugriffspfad der zum Typ des registerexkurrenten artikeladkurrent orientierten Zugriffspfades (*sensu* Wiegand 2007: 191ff) gehört und der beim Lemma **beruhigen** von wa_1 (vgl. Abb. 2) endet. Der Registereintrag re_4 aus (1), nämlich „Bitz s. bisschen, Kerngehäuse“ ist mehrfach außenorientiert. Ein mehrfach außenorientierter Registereintrag ist immer polydirektional. Die beiden in re_4 genannten Außenadressen sind in der gleichen äußeren Zugriffsstruktur zu finden, nämlich in der makrostrukturellen Hauptzugriffsstruktur von Seibicke (1989). Daher gehört re_4 zu den homogen außenorientierten Registereinträgen: Alle registerexkurrenten Zugriffspfade enden bei der gleichen äußeren Zugriffsstruktur. Der Registereintrag re_{44} aus (4) mit dem Registeringang „**Pejorativbildung**“ ist dagegen heterogen außenorientiert. Denn die Verweisadressenträger, die zu den Außenadressen |Schimpfwort| und |Spottname| gehören, sind Elemente der alphabetischen Registerzugriffsstruktur und damit Register-

eingänge, während alle anderen im gleichen Registereintrag genannten Verweisadressen von |Dicherling| bis |Wühler| als Verweisadressenträger ein Lemma aufweisen; demgemäß ist der Registereintrag re_{44} register- und lemmatisch-orientiert. Der Registereintrag re_{41} = „Paracelsus ↑ *Begriffs- und Bedeutungsprägung*“ ist einfach registerorientiert. Der Registereintrag r_{44} aus (4) in Abb. 1 ist mehrfach registerorientiert. Der Registereintrag re_{97} = „Leitelementträger 64* in (10) in Abb. 2 ist einfach umtextorientiert. Der Registereintrag re_{96} = „Leitelement 32*, 64*“ in (10) ist homogen außenorientiert, und zwar mehrfach umtextorientiert: Die beiden Hauptadressen |32*| und |64*| liegen in der Seitenzahlzugriffsstruktur zum Umtext „Lexikographisch-historische Einführung“. Der Registereintrag re_{104} = „Wörterbuchpflege s. Wörterbuchpflege, lexi-

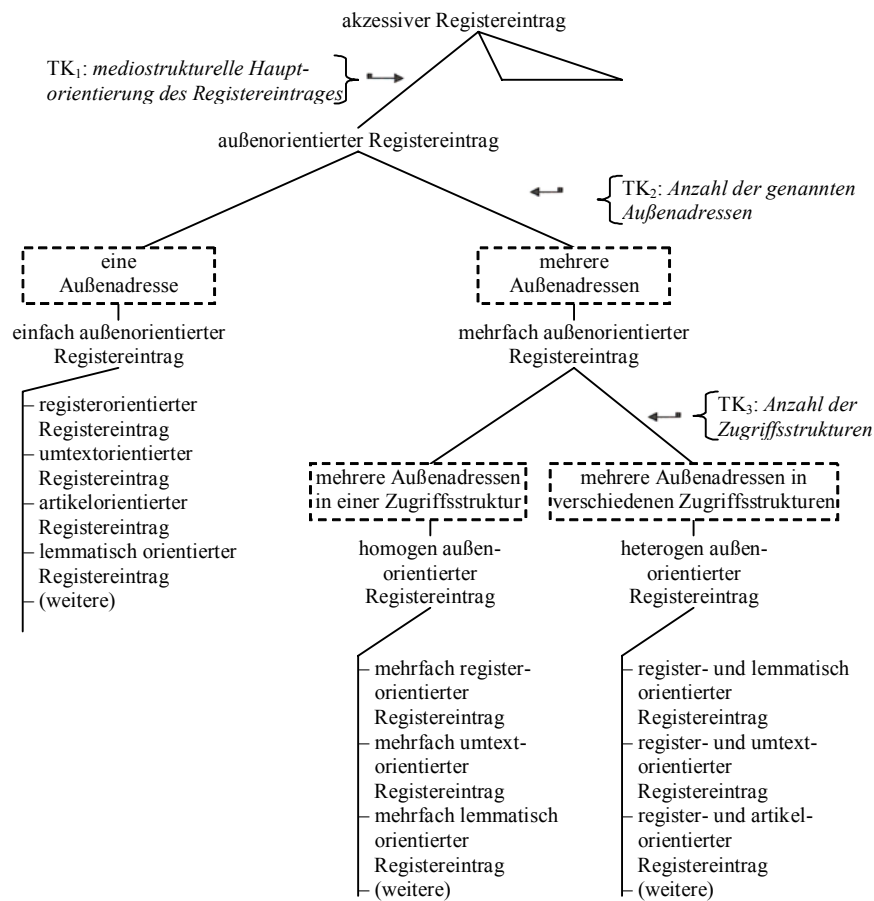


Abb. 4: Partiiell erweitert kommentierter Typologiegraph zum zweiten Typologieausschnitt aus einer Typologie von Registereinträgen in Printwörterbüchern nach der Art der mediostrukturellen Orientierung

kographische“ aus (10) in Abb. 2 ist einfach registerorientiert. Die Ausführungen zur mediostrukturellen Orientierung von Registereinträgen werden nachfolgend in einem Typologieausschnitt zusammengefasst.

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 4 gehören, seien die beiden Folgenden genannt:

(D-3: einfach außenorientierter Registereintrag)

Ein einfach außenorientierter Registereintrag ist ein einfacher Registereintrag, mit dessen Registerverweisangabe gerade eine Außenadresse genannt wird.

(D-4: heterogen außenorientierter Registereintrag)

Ein heterogen außenorientierter Registereintrag ist ein erweiterter Registereintrag, mit dessen Registerangaben mindestens zwei Außenadressen genannt werden, die in verschiedenen äußeren Zugriffstrukturen liegen.

2.1.2 Strukturen von Registereinträgen

Wie die anderen akzessiven Einträge, die aus Angaben bestehen, so weisen auch die Registereinträge — außer den vollständig reduzierten — konkrete hierarchische Strukturen auf, und zwar:

- konkrete hierarchische registerinterne Eintragskonstituentenstrukturen (kurz: registerinterne Konstituentenstrukturen)
- konkrete hierarchische registerinterne Eintragsmikrostrukturen (kurz: registerinterne Mikrostrukturen) und
- konkrete hierarchische registerinterne Eintragsangabenstrukturen (kurz: registerinterne Angabenstrukturen).

Zu allen genannten Strukturen gehören isomorphe abstrakte hierarchische Strukturen. Weiterhin weisen Registereinträge konkrete nichthierarchische Adressierungsstrukturen auf, zu denen isomorphe abstrakte angegeben werden können. Im Folgenden werden Beispiele für Registereintragsstrukturen behandelt.

Die meisten Strukturen von Registereinträgen sind wenig komplex. Zahlreiche Register weisen nur dann registerinterne Konstituentenstrukturen auf, wenn die Segmentationskonvention in Kraft gesetzt ist, dass registerinterne Leerzeichen (oder: Blanks) als nichttypographische registerinterne Mikrostrukturanzeiger gelten, so dass in der konkreten registerinternen Konstituentenstruktur Mikrostrukturanzeigerblanks (${}_i\text{MiSB}_i$) angesetzt werden können (wie in Abb. 5 (1)). Alle Registereinträge im Registerausschnitt (6) in Abb. 1 weisen z.B. nur dann eine konkrete hierarchische registerinterne Konstituentenstruktur auf, wenn der Blank zwischen Registeringang und numerischer Register-

angabe als Mikrostrukturanzeigerblank gilt. Die konkrete hierarchische registerinterne Konstituenten- und Mikrostruktur von re_{53} aus (6), dem Registereintrag „backen 111“, sind in Abb. 5 dargestellt.

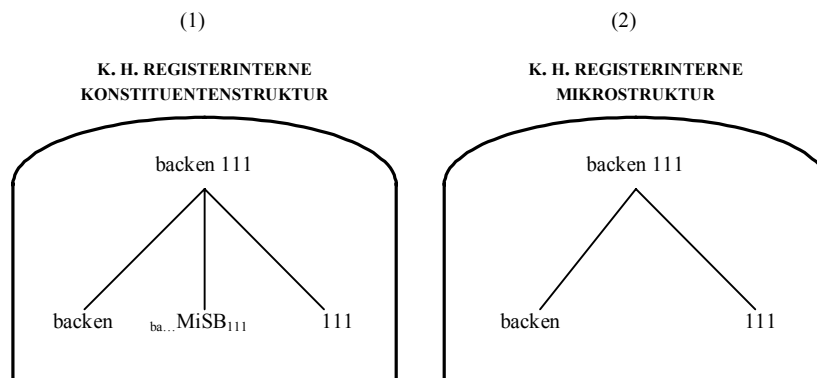


Abb. 5: Kommentierter Strukturgraph zur konkreten hierarchischen registerinternen Konstituentenstruktur des Registereintrags re_{53} aus (6) in Abb. 1 (1) und zur konkreten hierarchischen registerinternen Mikrostruktur von re_{53} (2)

Methodisch gesehen, sind die registerinternen hierarchischen Strukturen auf die gleiche Weise erhältlich wie die hierarchischen Artikelstrukturen. Dies sei im Folgenden für die registerinternen erweiterten Mikrostrukturen exemplarisch anhand des erweiterten vollständigen Registereintrages r_4 = „Bitz s. bisschen, Kerngehäuse“ aus (1) in Abb. 1 demonstriert. Zuerst wird r_4 mittels einer Anwendung der Methode der nichtexhaustiv funktionalpositionalen Segmentation (*sensu* Wiegand 2005a: 217ff) so segmentiert, dass folgende sechs Registereintragskonstituenten gegeben sind:

- (a) „Bitz“, der Registereingang (REg),
- (b) „s. bisschen, Kerngehäuse“, die nichtelementare registerinterne Verweisangabe (VerwA.ri),
- (c) „s.“, die registerinterne verdichtete Verweisbeziehungsangabe (v.VerwBA.ri),
- (d) „bißchen, Kerngehäuse“, die registerinterne homosegmentäre Außenadressenangabe (AuAdA.²ri),
- (e) „bißchen“, die registerinterne elementare Außenadressenangabe (AuAdA.ri),
- (f) „Kerngehäuse“, die registerinterne elementare Außenadressenangabe (AuAdA.ri).

Im Folgenden benutzen wir die Buchstaben a — f als Individuennamen für die Registereintragskonstituenten, re_4 als Individuennamen für den ganzen Registereintrag und bilden die strukturtragende Menge (oder: die Trägermenge) für

die konkrete (k) hierarchische registerinterne erweiterte Mikrostruktur von re_4 ; sie heie $M_T^k(re_4)$ und kann wie folgt angegeben werden: $M_T^k(re_4) = \{a, b, c, d, e, f, re_4\}$.

Auf der Menge $M_T^k(re_4)$ werden nun zwei ordnungsstrukturprgende Relationen definiert, und zwar die Folgenden:

- (1) Eine zweistellige Relation vom Typ der partitiven Relation (oder: Teil-Ganzes-Relation) — sie heie $R_{part}(re_4)$ — die reflexiv, antisymmetrisch und transitiv ist. $R_{part}(re_4)$ ist eine Teilmenge des einfachen Kreuzproduktes $M_T^k(re_4) \times M_T^k(re_4)$. Fr das einfache Kreuzprodukt einer Menge M mit sich selbst wird im Folgenden stets M^2 geschrieben; also gilt: $R_{part}(re_4) \subseteq (M_T^k(re_4))^2$. Der Relationsterm (= RT_1), der zu $R_{part}(re_4)$ gehrt, lautet: $RT_1 = x$ ist eine *Registereintragskonstituente* von y , mit x als Variable fr *Registereintragskonstituenten* und y als Variable fr *Registereintrge*, wie z.B. re_4 . RT_1 spezifiziert aus dem Kreuzprodukt $(M_T^k(re_4))^2$ genau die geordneten Paare, die zu $R_{part}(re_4)$ gehren, z.B. die Paare: (a, re_4) und (f, re_4) .
- (2) Weiterhin wird auf $M_T^k(re_4)$ ein zweistellige Relation vom Typ der Przedenzrelation (oder: Vorgnger-Nachfolger-Relation) — sie heie $R_p(re_4)$ — definiert; $R_p(re_4)$ ist irreflexiv (damit asymmetrisch) und transitiv und zu ihr gehrt der Relationsterm $RT_2 = x$ geht voraus y mit x und y als Variablen fr *Registereintragskonstituenten*. Es gelten: $R_p(re_4) \subseteq (M_T^k(re_4))^2$; $R_{part}(re_4) \cap R_p(re_4) = \emptyset$. Die Vereinigung von $R_{part}(re_4)$ und $R_p(re_4)$, nmlich $R_{part}(re_4) \cup R_p(re_4)$, ordnet den Registereintrag re_4 vollstndig. Die registerinterne Eintragsmikrostruktur von re_4 ist mithin eine Ordnungsstruktur.

Im nchsten Schritt weisen wir dann die *Registereintragskonstituenten* ihren Klassen zu, und zwar mittels Aussagen der Form $\lambda \in \beta$ (zu lesen wie λ ist ein *Element von* β). Fr „ β “ werden die oben bereits eingefhrten Klassensymbole verwendet. Demgem ergeben sich folgende Aussagen:

- $a \in REg$,
- $b \in VerwA.ri$,
- $c \in VerwBA.ri$,
- $d \in AuAdA^2.ri$,
- $e \in AuAdA.ri$,
- $f \in AuAdA.ri$.

Als Klassensymbol fr die Klasse aller *Registereintrge* (eines Registers) verwenden wir RE und bilden nun folgende Trgermenge fr die abstrakte (a) hierarchische registerinterne erweiterte Mikrostruktur von re_4 .

$$M_T^a(re_4) = \{REg, VerwA.ri, v.VerwBA.ri, AuAdA^2.ri, AuAdA.ri, RE\}.$$

Auf $M_T^a(re_4)$ werden nun ebenfalls eine Relation vom Typ der partitiven Relation und eine vom Typ der Präzedenzrelation definiert; diese Operation verläuft analog zu der im Falle der konkreten Struktur und wird daher nicht näher beschrieben.

Wie bei Artikelmikrostrukturen, so erfolgt die Darstellung von registerinternen Eintragsmikrostrukturen mittels geordneter Baumgraphen. In Abb. 3 sind der Baumgraph für die konkrete und der für die abstrakte Struktur zu einem einfach kommentierten Strukturgraphen dadurch zusammengefasst, dass sie aufeinander abgebildet sind. In einem solchen Strukturgraphen sind von der konkreten hierarchischen Eintragsmikrostruktur nur die terminalen mikrostrukturellen Konstituenten direkt wahrnehmbar; die nichtterminalen lassen sich jedoch anhand der vollständig sichtbaren abstrakten Struktur systematisch erschließen.

A.H. REGISTERINTERNE ERWEITERTE EINTRAGSMIKROSTRUKTUR

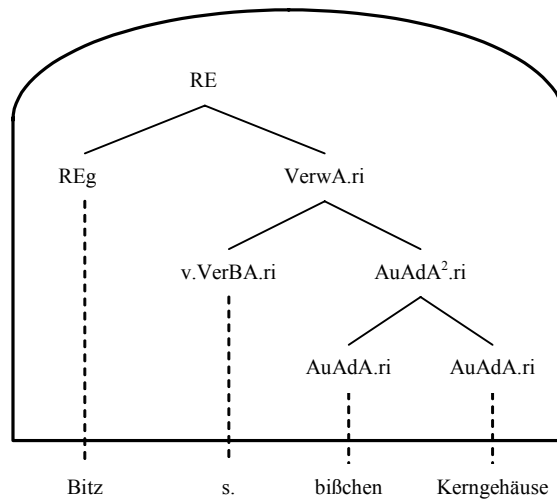


Abb. 6: Einfach kommentierter Strukturgraph zur abstrakten (und isomorphen konkreten) registerinternen erweiterten Mikrostruktur des Registereintrages re_4 in Abb. 1; „x---y“ bedeutet (von unten nach oben gelesen) soviel wie $x \in y$; „x—y“ bedeutet von (von unten nach oben gelesen) soviel wie x ist eine Registerintragskonstituente von y ; Abkürzungen: A = ABSTRAKT; H = HIERARCHISCH

In dem Register, aus dem (1) stammt, gibt es nur sehr wenige Registereinträge mit registerinternen homosegmentären Außenadressenangaben, und alle sind nur zweifach homosegmentär. Obwohl dazu in der Registervorbemerkung nichts gesagt wird, sind die elementaren Außenadressenangaben alphabetisch geordnet: „bißchen“ kommt alphabetisch vor „Kerngehäuse“. Auf die alphabe-

tische Anordnung der registerinternen Außenadressenangaben, die hier bewusst nicht berücksichtigt wurde, wird daher später zurückzukommen sein.

Im Folgenden betrachten wir den kondensierten Registereintrag r_{22} aus (2) in Abb. 1. Dieser weist die folgende Form auf:

Blaukabis	}	s. Kohl
Blaukappes		
Blaukraut		

Die exakte Analyse von r_{22} erfordert ein relativ elaboriertes Analyseinstrumentarium, das jedoch in der Heuristik, die zur Theorie der Wörterbuchform gehört, vollständig zur Verfügung steht. Zum besseren Verständnis betrachten wir zunächst vergleichend wa_2 in Abb. 2. Es handelt sich um einen verdichteten Verweisartikel; dieser kann verstanden werden als sei er im Zuge der äußeren Textverdichtung (*sensu* Wiegand 1998b: 28ff; vgl. auch Bustos Plaza und Wiegand 2006) aus zehn Verweisartikeln mit zehn gleichlautenden Verweiskommentaren „s. interim 1“, die alphabetisch unmittelbar aufeinander folgen, dadurch entstanden, das neun Mal der Verweiskommentar „s. interim 1“ getilgt wurde und die Verweislemmata zusammengerückt wurden. Statt die soeben nur grob skizzierte Methode der äußeren Textverdichtung anzuwenden, hätte man auch — wie in r_{22} — die Methode der akkoladenindizierten Textverdichtung anwenden können, dann hätte das Ergebnis (= wa'_2) wie folgt ausgesehen:

interimsabschied	}	s. interim 1
interimsaufwartung		
interimsbesitzer		
interimscondition		
interimseinteilung		
interimsunterhaltung		
interimsvermittlung		
interimsverordnung		
interimsweise		
interimswesen		

Der verdichtete Verweisartikel wa_2 in Abb. 2 ist mit wa'_2 funktionsgleich; wa_2 und wa'_2 sind jedoch strukturell verschieden. Während wa_2 eine (reine) Artikelkonstituentenstruktur aufweist und als deren Teilstruktur eine (reine) Artikelmikrostruktur, weist wa'_2 eine textuelle Struktur auf, die zu keinem der beiden Artikelstrukturtypen gehört. Mit Hinblick auf die Makrostruktur findet sich in wa_2 ein Lemmacluster vom Typ des alphabetischen Verweislemmaclusters, das aus zehn Verweislemmata besteht. Mit Hinsicht auf die Artikelmikrostruktur liegt eine homosegmentäre Lemmazeichengestaltangabe vor, die aus zehn Lemmazeichengestaltangaben besteht (LZGA¹⁰). In wa'_2 findet sich ein alphabetisches Verweislemmacluster mit den gleichen zehn Verweislemmata. Während jedoch in wa_2 auf der Menge aller clusterzugehörigen Lemmata nur

eine Relation vom Typ der striktinitialalphabetischen Relation definiert ist (mit dem Relationsterm $RT_3 = x \text{ geht } y \text{ striktinitialalphabetisch voraus}$, mit x und y als Variablen für Lemmata), ist in wa'_2 auf der gleichen Lemmatamenge zusätzlich eine textarchitektonische Relation vom Typ der *oberhalb*-Relation definiert (mit dem Relationsterm $RT_4 = x \text{ ist oberhalb von } y$, mit x und y als Variable für Lemmata).

Betrachtet man nun vergleichend re_{22} , dann gilt, dass re_{22} auch folgende Form aufweisen könnte:

$r'_{22} = \text{Blaukabis, Blaukappes, Blaukraut s. Kohl.}$

r_{22} und r'_{22} sind funktionsgleich aber strukturell verschieden. Während r'_{22} eine konkrete und isomorphe abstrakte registerinterne Mikrostruktur aufweist, die Teilstrukturen der registerinternen konkreten und isomorphen abstrakten registerinternen Eintragskonstituentenstruktur sind, weist r_{22} keine textuellen Strukturen dieses Typs auf. Vielmehr ist die Struktur, die r_{22} aufweist vom gleichen Typ wie die, die zu wa'_2 gehört.

Die Frage ist also nun: Zu welchem Strukturtyp gehört die textuelle Struktur von r_{22} ? Um diese Frage angemessen beantworten zu können, muss zuerst die Schweifklammer betrachtet werden. Es handelt sich im buchwissenschaftlichen Sinn um den hinteren Teil einer Akkolade. Eine Akkolade ist eine Schweifklammer „{...}“, mit der (beschriebene) Zeilen zu einer Menge von (beschriebenen) Zeilen, die unmittelbar übereinander liegen, zusammengeordnet werden. Dies kann, wie in r_{22} , auch nur mit dem hinteren Teil einer Akkolade „}“ erfolgen. „}“ besitzt eine linksgerichtete Klammeröffnung und eine rechte Klammerspitze. Im metalexikographischer Sprechweise handelt es sich um den hinteren Teil eines Zusammenordnungszeichens (hZOZ), das zugleich eine Textkondensierung anzeigt. „}“ ist damit ein nichttypographischer Kondensations- und Konstituentenstrukturanzeiger. Bei dem Registereintrag re_{22} in Abb. 1 (2) handelt es sich um einen Registereintrag mit einem Eingangskluster und akkoladenindizierter Textverdichtung. Die zugehörigen konkreten (und abstrakten isomorphen) Konstituentenstrukturen gehören zum Typ der hierarchischen zweiteiligen registerinternen Konstituentenstruktur mit vertikalarchitektonisch ausgebauter linker Teilstruktur. Konstituentenstrukturen dieses Typs weisen im Unterschied zu den allermeisten Artikelkonstituentenstrukturen als Teilstrukturen keine Mikrostrukturen auf. Der Grund dafür ist, dass in den Trägermengen von Mikrostrukturen (welcher Art auch immer) weder nichttypographische Strukturanzeiger noch Klassen solcher Strukturanzeiger enthalten sein dürfen. Eine angemessene Darstellung der textuellen Struktur von re_{22} ist jedoch ohne Berücksichtigung der Schweifklammer nicht möglich.

Im Folgenden wird die textuelle abstrakte Struktur, die von re_{22} zuerst in Abb. 7 dargestellt und daraufhin erläutert, wie sie erhältlich ist.

Im Folgenden wird grob erläutert, wie die in Abb. 7 dargestellte abstrakte Struktur, die zu re_{22} gehört, auf methodische Weise erhältlich ist. Zunächst wird

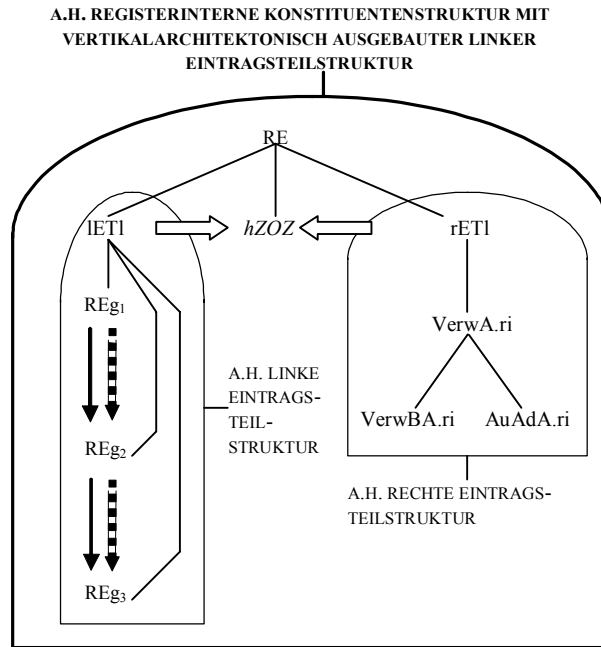
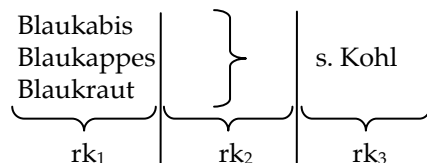


Abb. 7: Kommentierter mit architektonischen Komponenten angereicherter Strukturgraph zur abstrakten hierarchischen zweiteiligen registerinternen Konstituentenstruktur mit vertikalarchitektonisch ausgebauter linker Teilstruktur; „x—y“ bedeutet (von unten nach oben gelesen) soviel wie *x ist ein Teil von y*; „u $\overleftarrow{\text{---}}$ v“ bedeutet soviel wie *u ist oberhalb von v*; „r→s“ bedeutet soviel wie *r geht s alphabetisch voraus*; $a \Rightarrow b$ bedeutet soviel wie *a ist links von b*; $b \Leftarrow c$ bedeutet soviel wie *c ist rechts von b*; $a \Leftrightarrow b$ bedeutet soviel wie *a ist links von b und b ist rechts von a*. Abkürzungen: RE = Registereintrag; IETI = linker Eintragsteil; rETI = rechter Eintragsteil; hZOZ = hinteres Zusammenordnungszeichen

im ersten Schritt re_{22} durch eine Anwendung der Methode der erweiterten exhaustiven funktional-positionalen Segmentation segmentiert. Diese Methodenvariante lässt zu, dass eine Menge klassengleicher Angaben und Textsegmente, die im Text unmittelbar übereinander stehen, als genau eine nicht-elementare Textkonstituente gilt, so dass das folgende Segmentationsergebnis mit drei Registerkonstituenten gegeben ist (rk):



Im zweiten Schritt wird rk_1 in die drei Registereingänge $reg_1 = \text{„Blaukabis“}$, $reg_2 = \text{„Blaukappes“}$ und $reg_3 = \text{„Blaukraut“}$ segmentiert. Im dritten Schritt wird die registerinterne Außenadressenangabe „s. Kohl“ (= i) in „s.“, die verdichtete Verweisbeziehungsangabe (= j) und in „Kohl“, die Außenadressenangabe (= k), segmentiert. Dann werden im vierten Schritt die ermittelten Textsegmente wie folgt ihren Klassen zugeordnet: $rk_1; \in \text{IETI}$; $rk_2 \in \text{hZOZ}$: $reg_1 \in \text{Reg}_1$; $reg_2 \in \text{REg}_2$; $reg_3 \in \text{REg}_3$; $rk_3 \in \text{rETI}$; $i \in \text{VerwA.ri}$; $j \in \text{v.VerwBA.ri}$; $k \in \text{AuAdA.ri}$.

Im fünften Schritt wird die Trägermenge für die abstrakte hierarchische registerinterne Konstituentenstruktur mit vertikalarchitektonisch ausgebaute linker Eintragsteilstruktur — sie heiße $M_T^a(\text{re}_{22})$ — gebildet: $M_T^a(\text{re}_{22})$ kann wie folgt angegeben werden:

$$M_T^a(\text{re}_{22}) = \{\text{RE, IETI, hZOZ, rETI, REg}_1, \text{REg}_2, \text{REg}_3, \text{VerwA.ri, v.Verw. BA.ri, AuAdA.ri}\}.$$

In den nächsten sechs Schritten werden dann auf $M_T^a(\text{re}_{22})$ sechs strukturprägende Relationen definiert, und zwar die Folgenden:

- (1) Eine Relation vom Typ der partitiven Relation; sie heiße $R_{\text{part}}^a(\text{re}_{22})$; es gilt: $R_{\text{part}}^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *x ist ein Teil von y* (mit x und y als Variablen für Textsegmentklassen).
- (2) Eine Relation vom Typ der Präzedenzrelation; sie heiße $R_p^a(\text{re}_{22})$; es gilt: $R_p^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *x geht y voraus* (mit x und y als Variablen für Angabeklassen).
- (3) Eine Relation vom Typ der oberhalb-Relation; sie heiße $R_{\text{ob}}^a(\text{re}_{22})$; es gilt: $R_{\text{ob}}^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *u ist oberhalb von v* (mit u und v als Variablen für Klassen von Registereingängen).
- (4) Eine Relation vom Typ der links-von-Relation; sie heiße $R_{\text{lv}}^a(\text{re}_{22})$; es gilt: $R_{\text{lv}}^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *r ist links von v* (mit r und v als Variablen für Klassen von unmittelbaren Textkonstituenten des Registereintrags).
- (5) Eine Relation vom Typ der rechts-von-Relation; sie heiße $R_{\text{rv}}^a(\text{re}_{22})$; es gilt: $R_{\text{rv}}^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *v ist rechts von r* (mit v und r als Variablen für Klassen von unmittelbaren Textkonstituenten des Registereintrags. $R_{\text{lv}}^a(\text{re}_{22})$ und $R_{\text{rv}}^a(\text{re}_{22})$ sind konvers).
- (6) Eine Relation vom Typ der alphabetischen Relation; sie heiße $R_{\text{al}}^a(\text{re}_{22})$; es gilt: $R_{\text{al}}^a(\text{re}_{22}) \subseteq (M_T^a(\text{re}_{22}))^2$. Der Relationsterm lautet: *u geht v alphabetisch voraus* (mit u und v als Variablen für Klassen von Registereingängen).

Alle sechs auf $M_T^a(\text{re}_{22})$ definierten Relationen sind disjunkt.

In zahlreichen mehrfach erweiterten reduzierten Registereinträgen sind die reduzierten Registerverweisangaben alphabetisch geordnet, wie z.B. in re_{45} ,

dem Registereintrag mit dem Registerzugang „**Persisch**“ aus (4) in Abb. 1, oder den Registerinträgen re_{105} – re_{107} in (11) in Abb. 8.

(11)

Amerikanisches Guano Hängematte Kannibale Kanu Mokassin Orkan Savanne Tomahawk Totem Wigwam.

Amerikanische Pflanzen Ananas Chinin Curassao Kakao Kautschuk Kokain Mahagoni Mais Schokolade Tabak Tapioka Tomate Topinambur Yam Zigarre.

Amerikanische Tiere Jaguar Kakerlak Kolibri Kondor Lama N. Leguan Opossum Puma Skunk Tapir.

(12)

grassieren	splat	1571	V
gratis	lat	1558	Adv
Gratulant → gratulieren	(lat)	1743	No M
Gratulation → gratulieren	lat	1571	No F
gratulieren	lat	1563	V
gravieren	frz	1787	V
Gravität → gravitatisch	lat	1759	No F
gravitatisch	dt	1561	A
Grazie (frz, engl)	(lat)	1759	No F
graziös → Grazie	frz	1700	A
Grenadier	frz	1698	No M

(13)

Wahnsinnige, der 41,2 347,23
931,7
wahr 55,36 119,33 125,1
127,54 152,34 190,11 205,38
228,28 269,30 286,5 427,59
482,15 509,28 561,46 571,42
571,43 651,45 653,47 667,28
687,4 721,26 846,26 891,12
927,5 966,25 966,34 989,38
997,53 1008,40 1015,37
1063,3 1096,19
Wahre, das 63,19 65,56 71,39
96,52 111,25 119,33 168,48
174,43 176,3 177,63 192,22
192,26 192,34 192,49 267,50
269,31 399,12 413,40 418,32
423,60 425,31 426,64 428,36
544,7 547,28 549,20 551,5
573,18 634,49 658,8 670,42
680,51 681,36 686,48 711,60
715,44 746,50 749,15 787,19
796,14 810,45 827,6 914,2
1004,40 1045,19 1047,37
1061,56 1062,58 1065,13
1070,33
Wahre, der 668,47

(14)

abwehren aab-were
abweiden aab-etze
~, **mühsam** nägle²
~ **lassen** aab-etze
abweisen aab-wise
~ (**Bitte**) ab-schläà
~, **barsch** aab-pütze
~, **elegant** abschüfele,
abschüssele
abweisend üwéegg

Abb. 8: Registerausschnitt (11) aus Kluge (1943); (11) = re_{105} – re_{107} ; Registerausschnitt (12) aus Kirkness (1988); (12) = re_{108} – re_{118} . Registerausschnitt (13) aus Dobel (1972); (13) = re_{119} – re_{123} . Registerausschnitt (14) aus Bratschi und Trüb (1991); (14) = re_{124} – re_{132} .

Auch in mehrfach erweiterten vollständigen Registerinträgen, wie z.B. re_{82} in (8) in Abb. 2 mit dem Registerzugang „Entlehnung“, können die vollständigen Registerangaben alphabetisch geordnet sein. Die Registerinträge re_{82} und re_{105} gehören zum Typ des Registerintrages mit alphabetischer innerer Zugriffsstruktur. Registerinträge dieses Typs weisen eine konkrete hierarchische Mikrostruktur mit alphabetischer innerer Zugriffsstruktur auf, zu der eine isomorphe abstrakte angegeben werden kann. In Abb. 9 sind diese dargestellt.

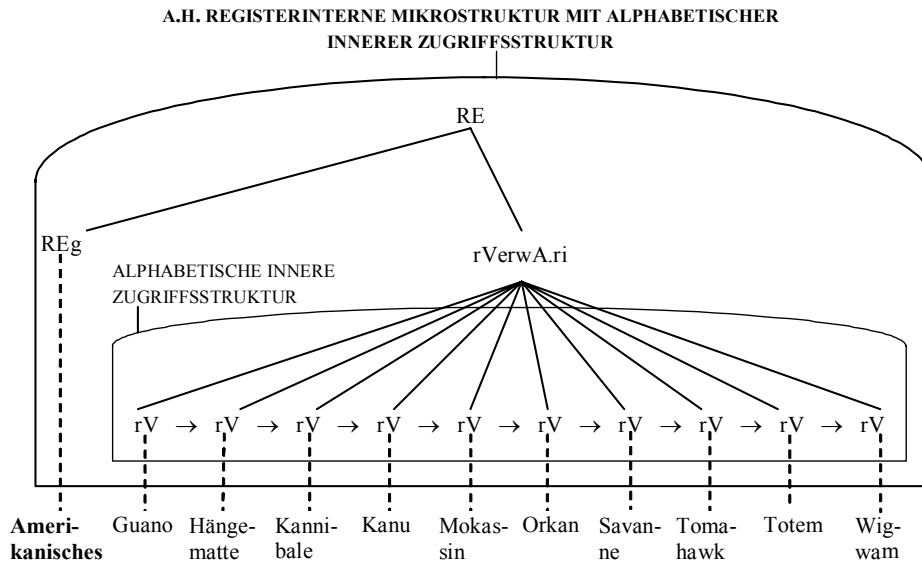


Abb. 9: Kommentierter Strukturgraph zur konkreten (und isomorphen abstrakten) hierarchischen registerinternen Mikrostruktur mit alphabetischer innerer Zugriffsstruktur, die re_{105} aus (11) in Abb. 8 aufweist. *Abkürzungen:* $rV = rVerwA.ri =$ registerinterne reduzierte Verweisangabe; „ $x \rightarrow y$ “ bedeutet soviel wie *x geht y striktinitialalphabetisch voraus*

Auf der Trägermenge einer hierarchischen registerinternen Mikrostruktur mit alphabetischer innerer Zugriffsstruktur ist neben einer Relation vom Typ der partitiven Relation und einer Relation vom Typ der Präzedenzrelation eine Relation vom Typ der striktinitialalphabetischen Relation definiert mit dem Relationsterm *x geht y striktinitialalphabetisch voraus* (mit *x* und *y* als Variablen für registerinterne reduzierte Verweisangaben). Die Präzedenzrelation wird benötigt, weil der Registereingang (in $re_{105} =$ **Amerikanisches**) der homosegmentären registerinternen reduzierten Verweisangabe nicht striktinitialalphabetisch vorausgeht; vielmehr geht er ihr voraus.

Im Registerausschnitt (13) in Abb. 8 weisen alle erweiterten reduzierten Registerinträge eine numerische innere Zugriffsstruktur auf. Die numerischen reduzierten Registerverweisangaben (z.B. „55, 36“) sind geordnete Paare, die aus einer registerinternen Spaltenzahlangabe („55“) und einer registerinternen Zeilenanzahlangabe („36“) bestehen. Spaltenzahlangaben heißen auch *Spaltenangaben* und Zeilenanzahlangaben auch *Zeilenangaben*. Die konkrete (und isomorphe abstrakte) hierarchische registerinterne Mikrostruktur mit numerischer innerer Zugriffsstruktur, die re_{119} aus (13) in Abb. 8 aufweist, ist in Abb. 10 dargestellt.

Auf den Trägermengen von registerinternen Mikrostrukturen mit numerischen inneren Zugriffsstrukturen sind neben partitiven und Präzedenzrelationen *kleiner-als*-Relationen definiert. Registerinträge mit numerischen redu-

zierten Verweisangaben, die als Teilangaben eine Spaltenangabe und eine Zeilenangabe aufweisen, funktionieren für den Benutzer dann, wenn das Wörterbuch neben einer äußeren Spaltenzahlzugriffsstruktur, auf jeder Wörterbuchseite nachgeordnete Zeilenanzahlzugriffsstrukturen aufweist. Dies ist in Dobel (1972) der Fall: Zu jeder der 1144 Spalten gibt es genau eine Zeilenanzahlzugriffsstruktur $\langle 10, 20, 30, 40, 50, 60 \rangle$, auf deren numerischen Zugriffstextelementen eine *oberhalb*-Relation definiert ist, so dass auf jedem Außen- und Innensteg der Wörterbuchseite eine gleichartige senkrechte Zeilenanzahlzugriffsstruktur verläuft, die dem Benutzer einen zeilengenauen Zugriff ermöglicht.

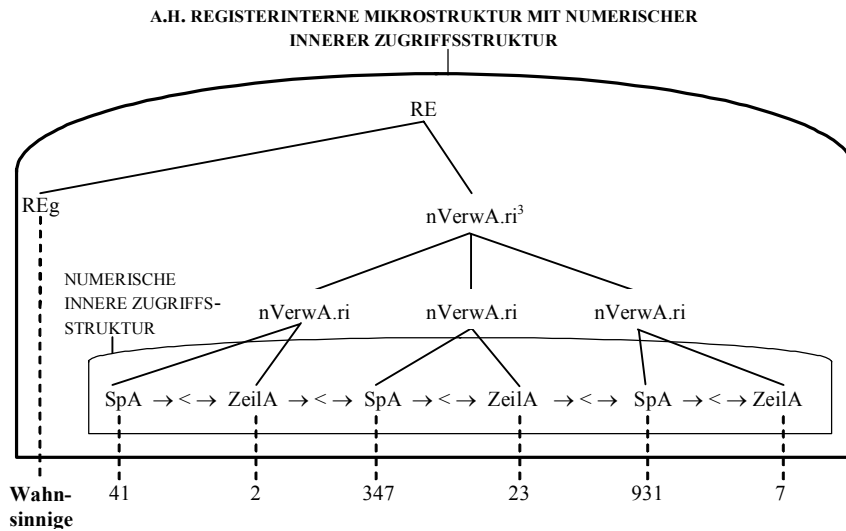


Abb. 10: Kommentierter Strukturgraph zur konkreten (und isomorphen abstrakten) registerinternen Mikrostruktur mit numerischer innerer Zugriffsstruktur, die re_{119} aus (13) in Abb. 8 aufweist: *Abkürzungen:* SpA = Spaltenangabe; ZeilA = Zeilenangabe; nVerwA.ri = registerinterne numerische Verweisangabe; „ $x \rightarrow < \rightarrow y$ “ bedeutet soviel wie *x ist kleiner als y*

Im Folgenden betrachten wir die Registerinträge in (12). Der Registerausschnitt (12) stammt aus dem alphabetischen Register von Kirkness (1988). Das zugehörige mehrbändige Deutsche Fremdwörterbuch ist nicht striktalphabetisch geordnet, so dass nicht alle Lemmata über die Hauptzugriffsstruktur erreichbar sind. Diejenigen Lemmata, auf die nicht direkt über die Hauptzugriffsstruktur zugegriffen werden kann, werden mit den Registereingängen genannt, auf die eine vollständige Verweisangabe folgt: z.B. „Gratulant \rightarrow gratulieren“ in re_{110} in (12) in Abb. 8. Alle Registerinträge aus (12), in denen auf den Registerzugang keine vollständige Verweisangabe folgt, also z.B. „grassieren“ und „gratis“ können als verweisvermittelnde Registereingänge gelten. In

Abb. 11 wird die konkrete (und isomorphe abstrakte) hierarchische Registermikrostruktur von re_{117} aus (12) in Abb. 8 dargestellt.

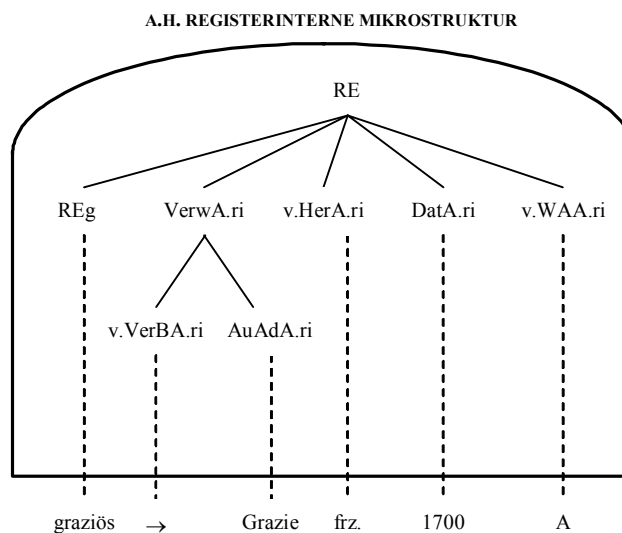


Abb. 11: Kommentierter Strukturgraph zur konkreten (und isomorphen abstrakten) registerinternen Mikrostruktur, die re_{117} aus (12) in Abb. 8 aufweist. *Abkürzungen:* v.HerA.ri = verdichtete registerinterne Herkunftsangabe; DatA.ri = registerinterne Datierungsangabe; v.WAA.ri = verdichtete registerinterne Wortangabe; A = Adjektiv; frz = französisch

Im Registerausschnitt (14) in Abb. 8 sind die Registereinträge, die mit einer Tilde beginnen, wie z.B. re_{130} = „~, **barsch** aab-pütze“, Registersubeinträge mit Registersubeingängen. Ein Benutzer-in-actu muss, wenn er den Registersubeintrag (oder: Registeruntereintrag) verstehen möchte, die Tilde durch den vorausgehenden Registerzugang ersetzen.

Wie in kondensierten Wörterbuchartikeln oder auch in anderen lexikographischen akzessiven Einträgen, die kondensiert sind und Angaben aufweisen, so treten auch in Registereinträgen Angabenadressierungsbeziehungen auf, in denen Registerangaben zu ihren Bezugsadressen stehen. Bei der Informationsgewinnung muss der Benutzer die adressierten Registerangaben auf ihre Bezugsadressen beziehen (vgl. Wiegand 2000; 2002a; 2006a).

Im Folgenden betrachten wir re_{124} aus (12) in Abb. 8. Dieser Registereintrag weist folgende adressierte Angaben auf:

- „splat“, die verdichtete registerinterne Herkunftsangabe (v.HerA.ri), mit der eine Abkürzung von *spätlateinisch* genannt wird.
- „1571“, die Datierungsangabe (DatA.ri)

- „V“, die verdichtete Wortangabe (v.WAA.ri), mit der eine Abkürzung vom *Verb* genannt wird.

Alle drei Angaben sind an den Registerzugang „grassieren“ linksadressiert, so dass in re_{124} die Adressierungskonstellation der vollständigen Registerzugangsadressierung vorliegt. Zu allen akzessiven Einträgen, in denen adressierte Angaben auftreten, kann eine konkrete Adressierungsstruktur angegeben werden sowie eine zugehörige isomorphe abstrakte.

Im Folgenden wird die abstrakte Adressierungsstruktur angegeben, die zu re_{124} gehört. Dazu wird zunächst eine Trägermenge für die abstrakte Registerzugangsadressierungsstruktur — sie heiße $M_{REAdS}^a(re_{124})$ — gebildet; diese muss die Klassen aller adressierten Registerangaben enthalten sowie die Klasse der bezugsadressentragenden Registerzugänge. $M_{REAdS}^a(re_{124})$ kann demgemäß wie folgt angegeben werden:

$$M_{REAdS}^a(re_{224}) = \{REg, v.HerA.ri, DatA.ri, v.WAA.ri\}.$$

Auf dieser Trägermenge wird eine zweistellige asymmetrische und irreflexive Relation — sie heiße $R_{REAdS}^a(re_{124})$ — definiert. Es gilt $R_{REAdS}^a(re_{124}) \subseteq M_{REAdS}^a(re_{124})^2$. Der Relationstyp ist festgelegt durch den Relationsterm x ist *eingangsintern adressiert an* y (mit x als Variable für adressierte Registerangaben und y als Variable für bezugsadressentragende Registerzugänge). Die strukturprägende Relation $R_{REAdS}^a(re_{124})$ induziert auf der Trägermenge $M_{REAdS}^a(re_{124})$ eine Partition $P(re_{124})$; das bedeutet, dass $R_{REAdS}^a(re_{124})$ die Trägermenge in zwei disjunkte Teilmengen zerlegt, nämlich in $M_{KadRA}^a(re_{124})$, die Menge der Klassen (K) der adressierten (ad) Registerangaben (RA), und in $M_{KbezREg}^a(re_{124})$, die Menge der Klassen (K) der bezugsadressentragenden (bez) Registerzugänge (REg). $P(re_{124})$ ist demgemäß wie folgt definiert: $P(re_{124}) = \{M_{KadRA}^a(re_{124}), M_{KbezREg}^a(re_{124})\}$. $M_{KadRA}^a(re_{124})$ ist der Vorbereich, $M_{KbezREg}^a(re_{124})$ ist der Nachbereich von $R_{REAdS}^a(re_{124})$. Die konkrete Adressierungsstruktur, die zu dem Registerzugang re_{124} gehört, ist erhältlich, indem man eine Trägermenge bildet, zu der alle adressierten Registerangaben von re_{124} sowie der bezugsadressentragende Registerzugang gehören und danach analog zur Bildung der abstrakten Struktur vorgeht. In Abb. 12 ist die abstrakte vollständig eingangsfokussierte Adressierungsstruktur (1) sowie die isomorphe konkrete vollständig eingangsfokussierte Adressierungsstruktur (2) von re_{124} dargestellt.

Im Registerzugang re_{124} aus (12) in Abb. 8, nämlich „Gravität → gravitatisch lat 1561 A“ tritt, wie auch in re_{110} , re_{111} und re_{117} aus (12), eine vollständige Verweisangabe auf, die aus zwei Teilangaben besteht, nämlich aus einer durch den Rechtspfeil „→“ realisierten Verweisbeziehungsangabe und der Außenadressenangabe „gravitatisch“. Die erste Teilangabe ist an die zweite adjazent rechtsadressiert, so dass in re_{124} die Adressierungskonstellation der partiellen Eingangsadressierung vorliegt. Zu Registerzugangseinträgen mit partieller Eingangsadressierung gehören Adressierungsstrukturen vom Typ der registerinternen partiell eingangsfokussierten Adressierungsstruktur.

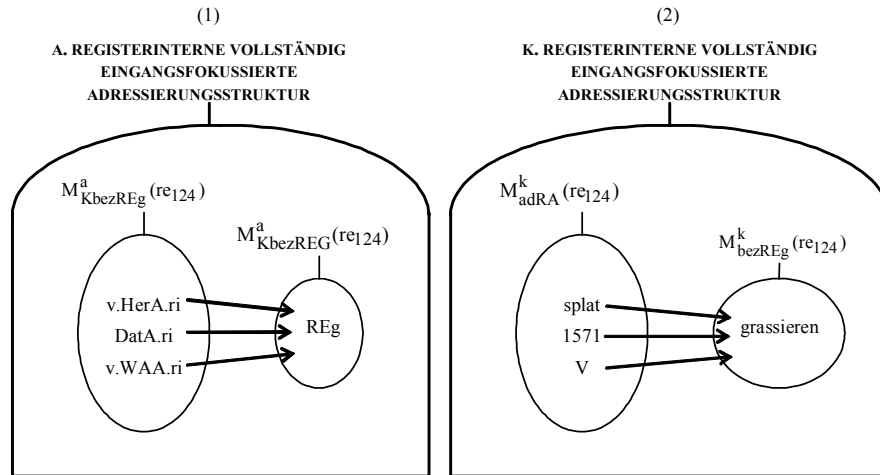


Abb. 12: Kommentierter Strukturgraph zur abstrakten (1) und zur isomorphen konkreten (2) registerinternen vollständig eingangsfokussierten Adressierungsstruktur, die re_{124} in (12) in Abb. 8 aufweist: *Abkürzungen:* A= ABSTRAKT; K= KONKRET; $M_{KadRA}^a(re_{124})$ = Menge der konkreten (k) adressierten (ad) Registerangaben (RA) von re_{124} ; $M_{KbezREG}^a(re_{124})$ = Menge der konkreten (k) bezugsadressentragende (bez) Registereingänge (REg) von re_{124}

Wenn man die registerinternen Mikrostrukturen eines Registerintrages um die zugehörigen Adressierungsstrukturen erweitert, sind die Angabenstrukturen dieses Registerintrages erhältlich. Die Erweiterung erfolgt so, dass auf den Trägermengen für beide Mikrostrukturen, neben den beiden mikrostrukturprägenden Relationen vom Typ der partitiven und der Präzedenzrelation, zusätzlich eine Relation vom Typ der Adressierungsrelation auf beiden Trägermengen definiert wird. Definiert man auf den Trägermengen der in Abb. 11 dargestellten konkreten (und isomorphen abstrakten) hierarchischen registerinternen Mikrostruktur von re_{117} zusätzlich eine Relation vom Typ der Adressierungsrelation, sind die konkrete (und isomorphe abstrakte) hierarchische registerinterne Angabenstruktur von re_{117} erhältlich.

Für deren Darstellung gibt es zwei Möglichkeiten. Entweder man reichert den Baumgraphen, mit dem die Mikrostruktur dargestellt wird um Pfeile an, die von einem Knoten, der eine adressierte Angabe repräsentiert, zu einem Knoten führt, der die bezugsadressentragende Angabe (und damit auch die Bezugsadresse) repräsentiert, oder man erweitert das Knotenetikett für eine adressierte Angabe um das Adressensymbol hinter dem Mittelpunkt, aus dem ersichtlich ist, welches Textsegment der Bezugsadressenträger ist (z.B. v.HerA.ri •REg). In Abb. 13, in der die konkrete (und isomorphe abstrakte) hierarchische registerinterne Angabenstruktur des Registerintrages re_{117} dargestellt ist, wurde das zuerst erläuterte Darstellungsverfahren gewählt.

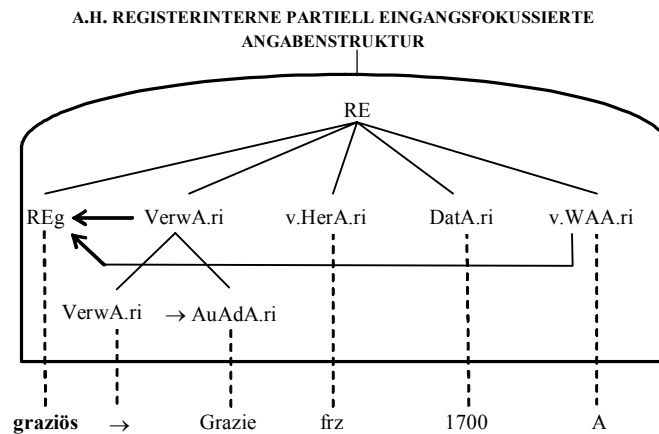


Abb. 13: Kommentierter Strukturgraph zur abstrakten (und konkreten isomorphen) hierarchischen registerinternen partiell eingangsfokussierten Angabenstruktur, die der Registereintrag re_{117} aus (12) in Abb. 8 aufweist; „ $x \rightarrow y$ “ bedeutet soviel wie *x ist an y adressiert*

Damit sind alle in Registereinträgen ausgeprägte Typen von Registereintragsstrukturen exemplarisch behandelt.

2.2. Typen von Registerzugriffsstrukturen und Typen von Registern

Jedes Wörterbuchregister weist eine Registerzugriffsstruktur auf. Registerzugriffsstrukturen sind immer äußere Zugriffsstrukturen (vgl. Wiegand 2008). Wörterbuchteile, die *Register* heißen, aber keine äußere Zugriffsstruktur aufweisen, gelten nicht als Register im Sinne der hier vorgelegten Theorie. Solche Fälle sind selten; ein Beispiel ist das so genannte onomasiologische Register in Scheuplein-Fritz, König, Krämer-Neubert und Wolf (1996, 183-204; vgl. (17) in Abb. 14). Auf dieses „Register“ ohne Registerzugriffsstruktur ist ein externer Zugriff nicht möglich. Wer etwas sucht, muss das Register durchlesen! Dass auch onomasiologische Register mit Registerzugriffsstrukturen angelegt werden können, zeigt das Sachregister in Heinzmann (1993).

Registerzugriffsstrukturen heißen wegen der angestrebten deutlichen Unterscheidung zu allen registerexternen Zugriffsstrukturen auch *registerinterne Zugriffsstrukturen*. Jedes Wörterbuchregister weist für seine Registerzugriffsstruktur einen direkten Registerzugriffsbereich (kurz: Zugriffsbereich) auf. Der direkte Registerzugriffsbereich ist die Menge aller akzessiven Registereinträge, die aus Registereingängen und Registerangaben bestehen, die meistens auf die Registereingänge folgen, wie man z.B. in den Registerausschnitten (1) bis (6) in Abb. 1 sehen kann. In seltenen Fällen können die Registerangaben auch den Registereingängen vorausgehen, wie z.B. in dem rückläufigen Register von Kirkness (1988; vgl. (18) in Abb. 14). In manchen Fällen können die Registeran-

gaben auf die linken Registereingänge folgen und zusätzlich den rechten Registereingängen einer zweiten Registerzugriffsstruktur vorausgehen, wie z.B. in (15) in Abb. 14, einem Ausschnitt aus einem Häufigkeitsregister mit dem Registertitel „Rangliste der Stichwörter“ in Erk (1972).

(15)

	D	F
tun	30	124
50 gehen	31	123
entscheiden	31	120
sogenannt- unterscheiden	32	117
erhalten	31	116
betrachten	30	108
auftreten	27	107
bringen	33	106
stellen	30	102
erreichen	28	101
60 nehmen	30	
gewinnen	27	98
erklären	29	97
halten	31	96

(16)

Ländername	Staatsbezeichnung
Südafrika	Republik Südafrika 232
Sudan	Republik Sudan 233
Südkorea	↑ Korea/Republik Korea 141
Suriname	Republik Suriname 234
Swasiland	Königreich Swasiland 235
Syrien	Syrische Arabische Republik 236
Tansania	Vereinigte Republik Tansania 237
Thailand	Königreich Thailand 238
Togo	Togolesische Republik 240
Tonga	Königreich Tonga 241

(17)

Wärme-, Lichtquelle

Trockenstangen am Ofen: Deise (3);

Fackeldeise; Herdstange; Ofendeise;

Ofengestänge; Rick (2); Spändeise

Kerze: Golllicht; Kienlicht; Licht

Docht: Dolich

Brenzellig riechen: brandeln; brandig;

räuchern; schmecken, nach Rauch; stin-

ken (2)

Feuer / Kerze anzünden: anstecken;

Feuer machen

Ofen anschüren: anstecken

Feuer im Ofen / Herd erhalten: anlegen

(2); hineinlegen

Elektrisches Licht: anbrennen (2);

andrehen (2); anknipsen

Essen und Trinken

Frühstuck: Kaffeetrinken (1); Morgen-
kaffee

Zwischenmahlzeit am Vormittag: Neu-
nuhr

Zwischenmahlzeit am Nachmittag:

Kaffeetrinken (2); Vieruhr (2)

(18)

No M	1689	dt	Summarist
No M	1718	dt	Seminarist
No M	1884	dt	Utilitarist
No M	1927	dt	Unitarist
No M	1775	frz	Äquibrist
			Equilibrist
No M	1460	(mlat)	Sacrist
No M	1750		Hypochondrist

(19)

11.Jh.	evangelisch 1	lat	A
	Quader 1	lat	No M F
1150	Synode	mlat,lat	No F
	Zepter	(lat,mlat) (griech)	No N M
1170	Sekte	mlat,lat	No F
1177	Vasall	(frz?) mlat	No M
1195	Revier 1	afrz	No N

(20)

Ägyptisch

Pharo (ägypt) 1728 No N

Algonkin

Wigwam (engl) (algonkin) 1697 No M

Arabisch

Algebra (lat) arab 1636 No F

Atlas I arab 15.Jh. No M

Gazelle (ital) arab 1611 No F

Harem arab 1779 No M

Mameluck (ital) arab 1519 No M

Moschee (frz,span) arab 1535 No F

Mumie (pers,arab) 1534 No F

Natron (span) arab 1530 No N

Ottomane 1 (arab) 1695 No F

Scheich arab 1654 No M

Talisman (span?,ital?) arab 1646 No M

Tarif (frz,ital) arab 1514 No M

Zenit (arab) 1490 No M

Ziffer (arab) 1399 No F

Babylonisch

Nimrod (babylon) 16. Jh. No M

Abb. 14: Registerausschnitt (15) aus dem nachspanninternen Häufigkeitsregister mit dem Registertitel „Rangliste der Stichwörter“; (15) = re₁₃₃–re₁₄₈. Registerausschnitt (16) aus Herzog und Hannes (1990); (16) = re₁₄₉–re₁₅₉. Registerausschnitt (17) aus Scheuplein-Fritz, König, Krämer-Neubert und Wolf (1996); (17): ohne akzessive Registereinträge. Registerausschnitt (18) aus Kirkness (1988); (18) = re₁₆₀–re₁₆₇; Registerausschnitt (19) aus dem chronologischen Register in Kirkness (1988); (19) = re₁₆₈–re₁₇₅. Registerausschnitt (20) aus dem Herkunftsregister in Kirkness (1988) (20) = re₁₇₆–re₁₉₆

Das Register, aus dem (15) in Abb. 14 stammt, weist eine linke (oder: linkslaufende) und eine rechte (oder: rechtslaufende) numerische Registerzugriffsstruktur auf. Von der linken sind nur die numerischen Registereingänge „50“ und „60“ zu sehen. Sie ist gedacht für den externen Zugriff mit Fragen vom Typ: WELCHE VERBEN HABEN DEN RANGLISTENPLATZ X–Y? Die rechte Zugriffsstruktur ist gedacht für Fragen nach der Frequenz im Korpus, die von einer Frequenzzahl ausgehen. Die Registerangaben sind die Verbangaben und die Distributionsangaben. Diese stehen somit vor und nach einem Registerzugang.

Register, in deren Registereinträgen nicht mindestens eine Registerverweisangabe auftritt, weisen nur einen direkten Zugriffsbereich auf. Solche Register gehören zum Typ des nichtmediostrukturellen Registers und ihre Registerzugriffsstruktur gehört entsprechend zum Typ der nichtmediostrukturellen Registerzugriffsstruktur. Beispiele für nichtmediostrukturelle Register sind das nachspanninterne „Register der Ehrengäste vom 1. bis 360. Mundart-Frühshoppen“ in Herwig (2005) und das so genannte Elementenregister in Karbelaschwili (2007) (vgl. dazu 4.).

Register, die durchgängig in ihren Registereinträgen mindestens eine Registerverweisangabe aufweisen, haben mindestens einen indirekten Registerzugriffsbereich. Dieser wird durch die mediostrukturelle Orientierung der Registerverweisangaben festgelegt. Die Register, zu denen die Registerausschnitte (1) und (2) sowie (4)–(7) in Abb. 1 gehören, weisen alle einen indirekten Zugriffsbereich auf. Sie gehören daher zum Typ des mediostrukturellen Registers. Mediostrukturelle Register haben eine mediostrukturelle Registerzugriffsstruktur. Sie heißen auch *Zugriffsregister*. Dieser Typenname ist durch den Sachverhalt motiviert, dass mit den Registerverweisangaben Außenadressen genannt werden, die dem Benutzer Zugriff auf die Zugriffssektoren des indirekten Zugriffsbereich eröffnen, weswegen Verweisadressen auch *Zugriffsadressen* heißen. In Wörterverzeichnissen sind die Zugriffssektoren die Wörterbuchartikel.

Mediostrukturelle Register können mehrere Zugriffsbereiche aufweisen. Wird in den Registereinträgen auf eine andere Textverbundkonstituente verwiesen und auch auf andere Registereinträge, wie z.B. in dem Sachregister von Paul (2002; vgl. (4) in Abb. 1 u. (8) in Abb. 2), dann hat das Register sich selbst auch als indirekten Zugriffsbereich, und die Registerzugriffsstruktur wird zur verweisdurchlässigen Zugriffsstruktur (i.S.v. Wiegand 2008). Mediostrukturelle Register mit gerade einem Registerzugriffsbereich heißen *monodirektionale Zugriffsregister*, solche mit mehreren Zugriffsbereichen heißen *polydirektionale Zugriffsregister*. Mehrere mediostrukturelle Register, die den gleichen Zugriffsbereich aufweisen, heißen *äquidirektionale Zugriffsregister*. Das NLO (1997) hat acht äquidirektionale Zugriffsregister: Alle sind Äquivalentregister und haben das Wörterverzeichnis als indirekten Registerzugriffsbereich. Das Deutsche Fremdwörterbuch hat fünf äquidirektionale Zugriffsregister: Ein alphabetisches (vgl. (12) in Abb. 8), ein rückläufiges, ein chronologisches (vgl. (19) in Abb. 14), ein Herkunftsregister (vgl. (20) in Abb. 14) und ein Wortklassenregister.

Der Typ des mediostrukturellen Registers weist zahlreiche Untertypen auf. Die wichtigsten Register in Printwörterbüchern sind die zentralen alphabetischen Zugriffsregister. Diese heißen *zentral*, weil ihr indirekter Registerzugriffsbereich das zentrale Wörterverzeichnis ist. Beispiele für solche Register sind die Wörterbuchregister, zu denen die Registerausschnitte (1) und (2) sowie (4)–(7) in Abb. 1 gehören, sowie die Register, zu denen (9) in Abb. 2 und (11)–(14) in Abb. 8 gehören. Weitere Beispiele für alphabetische zentrale Register sind das Register im Nachspann von Brunner und Moritz (1997), das Register im Nachspann von Augst (1984 und 2003), das nachspanninterne Wortregister in Riehme (1980) sowie das nachspanninterne Register in Schröder (1992). Der registervermittelte Zugriff, der durch alphabetische zentrale Register ermöglicht wird, kann über folgende äußere Zugriffsstrukturen erfolgen:

- über die alphabetische Hauptzugriffsstruktur: Mit den Registerverweisangaben werden lemmatische Verweisadressen genannt, wie z.B. in (11) in Abb. 8.

- über die Seitenzahlzugriffsstruktur: Mit den Registerverweisangaben werden Seitenzahlen als numerische Verweisadressen genannt, wie z.B. in (6) in Abb. 1.
- über die Seitenzahlzugriffsstruktur mit nachgeordneter Zeilenzahlzugriffsstruktur: Die Registerverweisangaben sind numerische Doppeladressenangaben, mit denen als Hauptadresse eine Seitenzahl und als Unteradresse eine Zeilenzahl genannt wird.
- über die Spaltenzahlzugriffsstruktur mit nachgeordneter Zeilenzahlzugriffsstruktur: Die Registerverweisangaben sind numerische Doppeladressenangaben, mit denen als Hauptadresse eine Spaltenzahl und als Unteradresse eine Zeilenzahl genannt wird, wie z.B. in (13) in Abb. 8.
- über die alphanumerische Kopfzeilenzugriffsstruktur in Sachgruppenwörterbüchern: Mit den Registerverweisangaben werden die Sachgruppennummer und der zugehörige Artikeltitle genannt, wie z.B. in (5) in Abb. 1.

Neben dem Typ des zentralen alphabetischen Zugriffsregisters lässt sich der Typ des peripheren alphabetischen Zugriffsregisters unterscheiden, der mehrere Untertypen aufweist. Register dieses Typs haben als indirekten Registerzugriffsbereich Umtexte oder Einschübe. Beispiele für periphere alphabetische Zugriffsregister sind: Das vortextinterne Register zum Umtext mit dem Umtexttitel „Einführung in die Terminologie“ in Kluge und Seebold (1995); weiterhin das vortextinterne Register mit dem Titel „Stichwortregister zur lexikographischen Einleitung“ im FWB-1 sowie das Register zur „Lexikographisch-historischen Einführung“ in Dornseiff (2004) und das „Register zum Vorwort“ in WDG-1. Ein peripheres alphabetisches Zugriffsregister zur „Systematischen Einführung“ wird sich auch im „Wörterbuch zur Lexikographie und Wörterbuchforschung“ (WLWF) finden (vgl. Wiegand 2003a) sowie in allen geplanten WSK-Bänden (vgl. Schierholz und Wiegand 2005). Alle genannten peripheren alphabetischen Zugriffsregister gehören zum Untertyp des umtextorientierten alphabetischen Zugriffsregisters.

Eine weitere typologische Unterscheidung ist gegeben mit dem Typ des orientierenden und dem des nichtorientierenden Zugriffsregisters. Ein orientierendes alphabetisches Zugriffsregister weist nicht nur Fundortdaten in der Form unterschiedlicher Registerverweisangaben auf, sondern auch andere Angaben, so dass angereicherte Registerinträge gegeben sind; oder es sind bifunktionale Angaben gegeben wie in den Äquivalentregistern. Entsprechend haben orientierende Register adressen- und inhaltsvermittelnde Registerzugriffsstrukturen. Meistens sind orientierende Zugriffsregister zentrale alphabetische Register. Der Registerausschnitt (7) in Abb. 1 stammt aus einem zentralen orientierenden alphabetischen Register; ebenso der Registerausschnitt (12) in Abb. 8 sowie (16) in Abb. 14. Die inhaltliche Orientierung, die ein Benutzer anhand eines orientierenden Zugriffsregisters erhält, kann er vor der Ausfüh-

zung einer Verweisbefolgungshandlung erschließen. Anders ausgedrückt heißt das: Ein orientierendes Zugriffsregister ermöglicht die Erschließung von Antworten auf bestimmte Suchfragen, ohne dass das zentrale Wörterverzeichnis oder eine andere Textverbundkonstituente registervermittelt konsultiert wird. Ein Benutzer, der z.B. lediglich wissen möchte, wie die offizielle Staatsbezeichnung von Syrien ist, erhält die Antwort anhand einer Registerangabe des Registers, aus dem der Ausschnitt (16) stammt. Zentrale orientierende alphabetische Zugriffsregister treten in unterschiedlichen Ausprägungen auf. Alle zweisprachigen Äquivalentregister, z.B. die in NLO (1997), gehören zu den zentralen orientierenden alphabetischen Zugriffsregistern. Für alle Äquivalentregister ist charakteristisch, dass mit ihrer Hilfe Antworten auf Suchfragen vom Typ WIE LAUTET DAS ÄQUIVALENT ZU EINEM AUSDRUCK DER SPRACHE X IN EINER SPRACHE Y? erhältlich sind.

Nichtorientierende Zugriffsregister, die auch *reine Zugriffsregister* heißen, weisen adressenvermittelnde Registerzugriffsstrukturen auf. Die Register, zu denen die Registerausschnitte (1), (2), (4) und (6) in Abb. 1 gehören, sind beispielsweise zentrale nichtorientierende alphabetische Zugriffsregister. Auch periphere Register können nichtorientierende Zugriffsregister sein. Beispielsweise ist das Register, aus dem (10) in Abb. 2 stammt, ein peripheres nichtorientierendes alphabetisches Zugriffsregister.

Die meisten Registerzugriffsstrukturen sind glatte äußere Zugriffsstrukturen, so dass auf der Trägermenge auch eine textarchitektonische *oberhalb*-Relation definiert ist; das gilt z.B. für die Register, aus denen die Registerausschnitte (1), (2) und (4)–(6) in Abb. 1 oder auch für die Register, aus denen (16) und (18) in Abb. 14 stammen. Neben den glatten Registerzugriffsstrukturen treten jedoch auch gestaffelte Registerzugriffsstrukturen auf (vgl. Wiegand 2008). Diese bestehen aus einer vorgeordneten Registerzugriffsstruktur, sowie n nachgeordneten Registerzugriffsstrukturen (mit $n > 1$). Bei gestaffelten äußeren Zugriffsstrukturen, heißt die vorgeordnete Zugriffsstruktur auch *primäre Zugriffsstruktur* und die nachgeordnete Zugriffsstruktur *sekundäre Zugriffsstruktur*. Die Registerzugriffsstruktur des Chronologischen Registers in Kirkness (1988) ist eine gestaffelte Registerzugriffsstruktur (vgl. (19) in Abb. 14). Die vorgeordnete Registerzugriffsstruktur ist alphanumerisch. Ihre Registereingänge sind entweder Jahreszahlen, wie z.B. „1150“ oder alphanumerische Ausdrücke der Form „11. Jh.“. Die nachgeordneten Registerzugriffsstrukturen sind alle glattalphabetisch. Die Registerzugriffsstruktur des Chronologischen Registers in Kirkness (1988) gehört daher zum Typ der gestaffelten alphanumerisch-chronologischen/alphabetischen Registerzugriffsstruktur. — Auch das Herkunftsregister in Kirkness (1988) hat eine gestaffelte Registerzugriffsstruktur (vgl. (20) in Abb. 14); Diese gehört zu einem anderen Untertyp, und zwar zum Typ der gestaffelten alphabetischen/alphabetischen Registerzugriffsstruktur. Sowohl die vorgeordneten Registereingänge, z.B. **Ägyptisch**, **Algonkin**, **Arabisch**, **Babylonisch** (die auch *Hauptregistereingänge* heißen), als auch nachgeordneten Registereingänge (die auch *Unterregistereingänge* heißen) sind alpha-

betisch geordnet, und zwar im Registerausschnitt (20) striktinitialalphabetisch. — Schließlich weist auch das Wortklassenregister in Kirkness (1988) eine gestaffelte Registerzugriffsstruktur auf (vgl. (21) in Abb. 15).

(21)			(22)		
Nomen als Bestimmungswort			Märchen		
Detail-	frz	1782	Medien		
Malefiz-	lat	15. Jh.	Meistergesang		
Pleinair-	(frz)	1889	Metapher		
Radio-	(lat)	1930	Metrik		
Sanität-		1814	Mimesis		
Statutar-	dt	1889	Minnesang		
Synthetik-	dt	1962	Mittelalterrezeption		
Tele-	dt	1956	Mittellateinische Literatur		
Teleskop-	(engl)	1898	Moderne		
Tropik-	(engl?) (lat,griech)	1778	Mündlichkeit/Schriftlichkeit		
Nomen als Grundwort			Mystik		
-technik	gellat (griech)	1711	Mythos		
(23)			(24)		
csatol	54		brevitas	0436	
csatolt tartományok	80		brisare	0440a*	
csökkentett járulékfizetés	338		brocchus	0440b	
csökkentett (összegű) nyugdíj	224		bruma	0441a	
csökkent munkaképesség	132, 339		{bruscum	0441b}	
csökkent munkaképességű	126, 131		brutalis (brutus)	0442	
csökkent munkaképességűek járadéka	297		bubalus (bufalus)	0449	
			bucca (buccula)	0428	
			buccula	0445, 1671	
			bucina	2735a	
			bufo	0450	
			bulga	0448	

Abb. 15: Registerausschnitt (21) aus dem Wortklassenregister in Kirkness (1988); (21) = re₁₉₇–re₂₁₃. Ausschnitt (22) aus dem „Verzeichnis der Artikel“ aus Brunner und Moritz (1997). Registerausschnitt (23) aus Ildikó (2005); (23) = re₂₁₄–re₂₂₀. Registerausschnitt (24) aus Kytzler und Redemund (2007); re₂₂₁–re₂₃₂.

Bei der äußeren Zugriffsstruktur des Wortklassenregisters in Kirkness (1988) handelt es sich ebenfalls um eine Registerzugriffsstruktur vom Typ der gestaffelten alphabetischen/alphabetischen Registerzugriffsstruktur. Mit den Hauptregistereingängen werden Termini für Wortklassen genannt und mit den Unterregistereingängen Ausdrücke, die zur jeweiligen Wortklasse gehören. Die Unterregistereingänge sind verweisvermittelnde Registereingänge.

In manchen Wörterbüchern finden sich alphabetische Listen, die lediglich aus Eingängen bestehen. Ein Beispiel ist die „Liste der Wortbildungselemente, die in der alphabetischen Stichwortliste erscheinen“ in Duden-10, aus der (3) in Abb. 1 stammt. Ein weiteres Beispiel ist das vorspanninterne alphabetische „Verzeichnis der Artikel“ in Brunner und Moritz (1997), aus dem (22) in Abb.

15 stammt. Schließlich sei als letztes Beispiel die nachspanninterne alphabetische Liste in Götz und Wellmann (2007) genannt, die mit „Definitionswortschatz“ überschrieben ist. Solche Listen finden sich in zahlreichen Lernerwörterbüchern.

Für solche Listen gibt es m.E. zwei Möglichkeiten für eine angemessene lexikographietheoretische Deutung. Entweder handelt es sich um ein reduziertes alphabetisches Register, also um ein Register ohne Registerangaben, oder es liegt ein rudimentäres lexikographisches Inhaltsverzeichnis vor, in dem die Nennung der Seitenzahlzugriffstruktur fehlt. Welche der beiden Möglichkeiten vorliegt, entscheiden die jeweiligen textuellen Gegebenheiten in dem Wörterbuch. Moritz und Brunner (1997) ist ein Fachwörterbuch, das relativ wenige (nämlich 144) umfangreiche Fachwörterbuchartikel aufweist. Dem vorspanninternen alphabetischen Artikelverzeichnis (vgl. die *M*-Strecke in (22) in Abb. 15) kann daher die genuine Funktion zugewiesen werden, dem Benutzer einen schnellen Überblick über die Fachwörterbuchartikel zu ermöglichen. Die gleiche Funktion hätte ein lexikographisches Inhaltsverzeichnis. Da in Moritz und Brunner (1997) ein nachspanninternes ausführliches alphabetisches Register literaturwissenschaftlicher Termini existiert, liegt in diesem Fall eine Deutung nahe, das Artikelverzeichnis, auch weil es alle Artikel nennt und damit einen Bezug zum Ganzen hat, als rudimentäres lexikographisches Inhaltsverzeichnis aufzufassen. Im übrigen gilt, dass der Status des Rudimentären nur strukturell gegeben ist. Funktional gesehen, besteht er nicht, da ein alphabetisch geordnetes rudimentäres lexikographisches Inhaltsverzeichnis einen lemmagenauen Zugriff gestattet.

Die „Liste der Wortbildungselemente, die in der alphabetischen Stichwortliste erscheinen“ (vgl. (3) in Abb. 1), die sich — nebenbei bemerkt — auch im Vorspann des Duden-DaF findet — lässt sich nicht als rudimentäres lexikographisches Inhaltsverzeichnis deuten; ihr fehlt der Bezug zum gesamten Textverbund. Mit einer alphabetischen Liste der als Lemmata angesetzten Wortbildungselemente (nämlich der Affixe und Affixoide) werden automatisch auch die zugehörigen Affix- und Affixoidartikel aufgelistet, so dass ein partielles Artikelverzeichnis gegeben ist. Dieses lässt sich als reduziertes Register auffassen, dessen Reduzierung darin besteht, dass auf die Registereingänge keine Registerangaben folgen. Die Registerinformation, die insgesamt erhältlich ist, besteht darin, dass ein Benutzer erfährt, welche Wortbildungsmittel als Lemmata angesetzt sind sowie für ein spezifisches Suchinteresse darin, ob ein bestimmtes Wortbildungsmittel als Lemma angesetzt ist oder nicht. — Auch die alphabetische Liste des Definitionswortschatzes kann als reduziertes alphabetisches Register aufgefasst werden. Register — so hatten wir gesehen — sind entweder mediostrukturell oder nichtmediostrukturelle Register. Daher entsteht die Frage, zu welchem der beiden Typen die reduzierten alphabetischen Register als Untertypen gehören. Angenommen ein Benutzer möchte wissen, ob das Wort *Basis* zum Definitionswortschatz von Götz und Wellmann (2007) gehört. Dann kann er mit dem Leitelement |Basis| eine Benutzungshandlung vom Typ der externen Registerzugriffshandlung ausführen. Wenn er den Re-

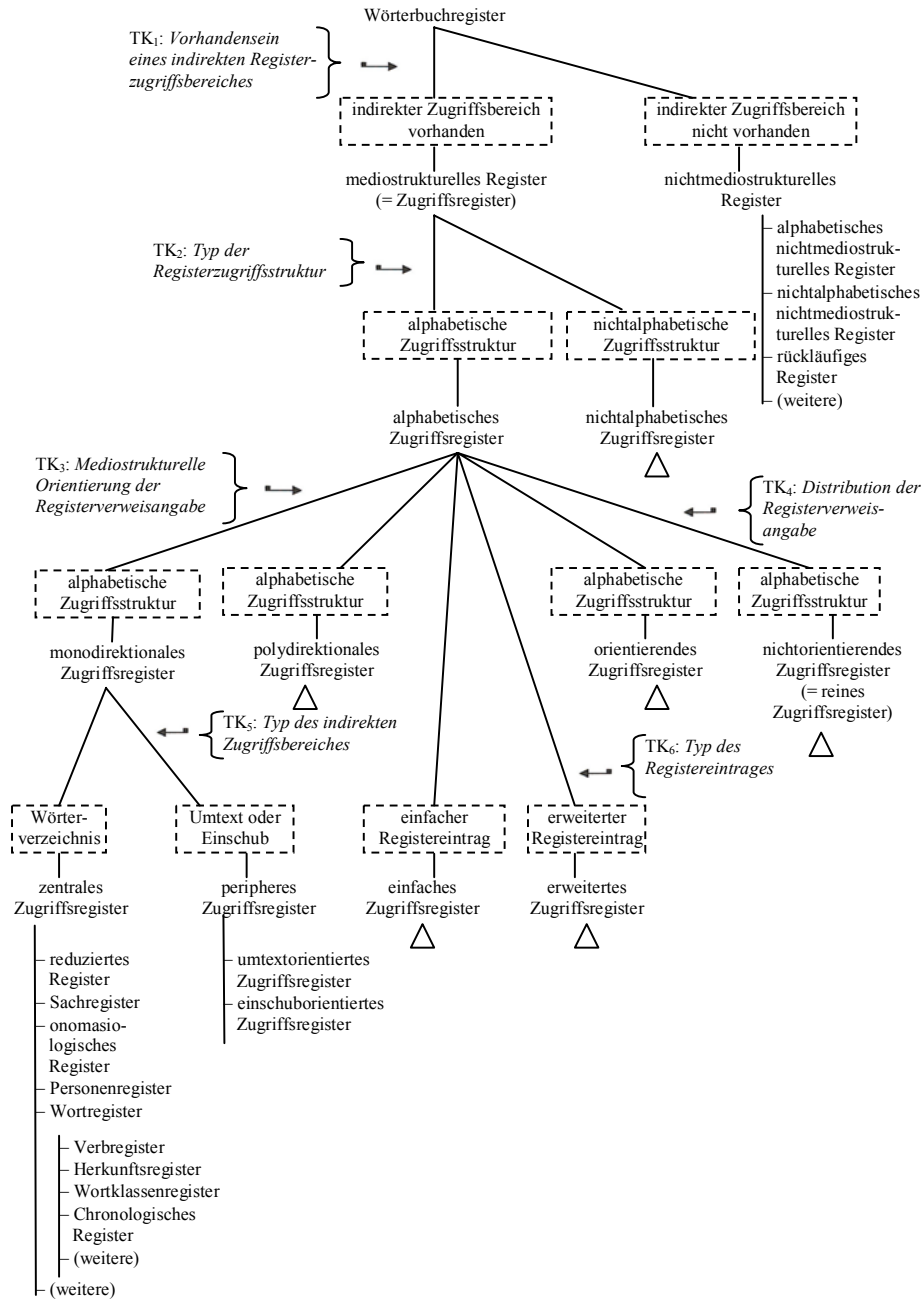


Abb. 16: Partiiell kommentierter Typologiegaph zu einem Ausschnitt aus einer Typologie von Wörterbuchregistern

gistereingang „Basis“ gefunden hat und möchte nun z.B. wissen, ob die Bedeutungsparaphrase zu *Basis*, die mit der Bedeutungsparaphrasenangabe im Wörterbuchartikel zu *Basis* genannt wird, nur mit Wörtern aus der Liste, 'Definitionswortschatz' formuliert ist, dann muss er zuerst im Wörterbuchartikel mit dem Lemma **Basis** die Bedeutungsparaphrasenangabe suchen. Da der Benutzer von dem reduzierten alphabetischen Register zum Lemma **Basis** gehen muss, lässt sich die Benutzungshandlung, die er ausführen muss, als registervermittelte externe Verweispflichtungshandlung auffassen. Diese Auffassung setzt voraus, dass die Registereingänge von reduzierten alphabetischen Registern, mit deren Registereingängen Lemmata erwähnt werden, als verweispflichtende Registereingänge gelten, so dass sich folgender Vorschlag ergibt: Reduzierte alphabetische Register, mit deren Registereingängen Lemmata erwähnt werden, gelten als mediostrukturelle Register; ihr indirekter Registerzugriffsbereich sind Wörterverzeichnisse. Dagegen gelten reduzierte alphabetische Register — wie z.B. das „Register der Ehrengäste vom 1. bis 360. Mundart-Frühshoppen“ in Herwig (2005) — als nichtmediostrukturelle Register; sie weisen nur einen reduzierten Registerzugriffsbereich auf, der aus der Menge aller Registereingänge besteht, mit denen keine Lemmata erwähnt werden.

Im Abb. 16 fassen wir zunächst die unterschiedenen Registertypen zu einem Typologieausschnitt zusammen.

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 16 gehören, seien die beiden Folgenden genannt:

(D-5: zentrales Zugriffsregister)

Ein zentrales Zugriffsregister ist ein alphabetisches monodirektionales Wörterbuchregister, dessen indirekter Zugriffsbereich das Wörterverzeichnis ist.

(D-6: erweitertes Register)

Ein erweitertes Register ist ein alphabetisches Zugriffsregister in einem Wörterbuch, das erweiterte Registerinträge aufweist.

Der Abschnitt sei beendet mit einem Ausschnitt aus einer Typologie von Registerzugriffsstrukturen (vgl. auch Wiegand 2008).

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 17 gehören, seien die beiden Folgenden genannt:

(D-7: alphabetische zentrale mediostrukturelle Registerzugriffsstruktur)

Eine alphabetische zentrale mediostrukturelle Registerzugriffsstruktur ist eine mediostrukturelle Zugriffsstruktur eines zentralen alphabetischen Zugriffsregisters, deren indirekter Registerzugriffsbereich das zentrale Wörterverzeichnis ist.

(D-8: alphabetische nichtmediostrukturelle Registerzugriffsstruktur)

Eine alphabetische nichtmediostrukturelle Registerzugriffsstruktur ist eine Registerzugriffsstruktur, deren Registereingänge alphabetisch geordnet sind und die keinen indirekten Registerzugriffsbereich aufweist.

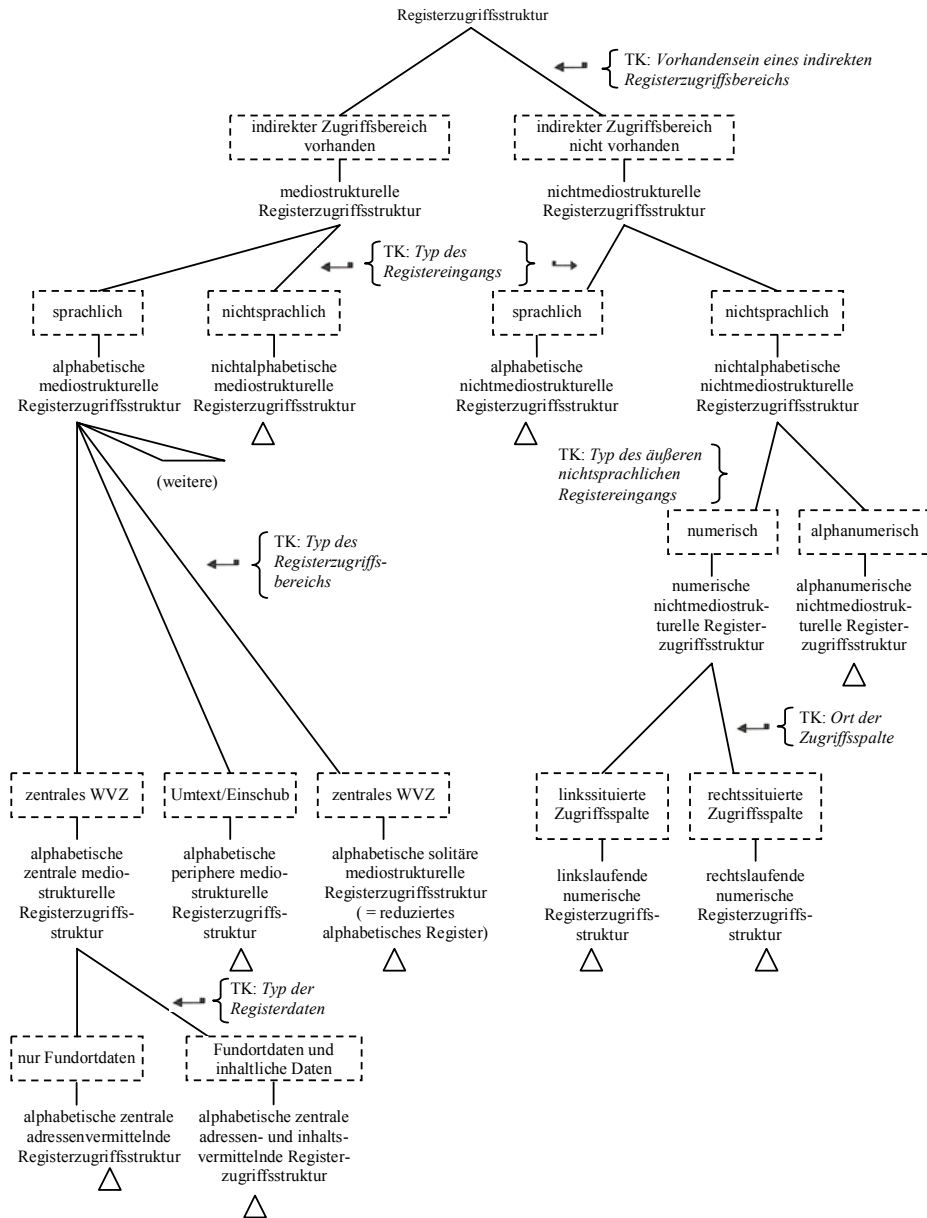


Abb. 17: Kommentierter Typologigraph zu einem Ausschnitt aus einer Typologie von Registerzugriffsstrukturen; Abkürzung: WVZ = Wörterverzeichnis

3. Registerfunktionen

Unter einer Registerfunktion wird eine Funktion (i.S.v. *Aufgabe*) verstanden, die ein Register im Rahmen des gesamten Printwörterbuchs hat. Die jeweilige Funktion eines Registers wird erstens wesentlich bestimmt von den Eigenschaften, die das Wörterbuch hat, zu dem das Register gehört und damit vom jeweiligen Wörterbuchtyp. Zweitens bestimmt der jeweilige Registertyp die Registerfunktion.

Zunächst gilt allgemein stets das Folgende: Da jedes Register eine Registerzugriffsstruktur aufweist und da alle Registerzugriffsstrukturen äußere Zugriffsstrukturen sind, leistet jedes Register einen Beitrag zur externen Datenakzessivität des Wörterbuchs, zu dem es gehört. Will man unterschiedliche Registerfunktionen bestimmen und näher charakterisieren, muss man daher den jeweiligen Beitrag eines Registers zur externen Datenakzessivität spezifizieren, und zwar unter Berücksichtigung des Wörterbuchtyps und des Registertyps.

Wir betrachten zunächst alphabetische zentrale Zugriffsregister, also solche monodirektionalen Register, deren indirekter Registerzugriffsbereich ein Wörterverzeichnis ist und beginnen mit alphabetischen zentralen Zugriffsregistern in so genannten onomasiologischen Wörterbüchern, die auch unter anderen Typennamen bekannt sind, wie z.B. *systematisches Wörterbuch* (vgl. z.B. Friederich 1966). Ein Ausschnitt aus dem alphabetischen zentralen Zugriffsregister im Dornseiff (2004) ist (5) in Abb. 1. Weitere Beispiele sind das „Alphabetische Register der Mundartwörter“ sowie das „Alphabetische Register hochdeutscher Stichwörter“ in dem Dialektwörterbuch nach Sachgruppen (Meng 1986). Solche Register, deren indirekter Zugriffsbereich ein Wörterverzeichnis ohne alphabetische makrostrukturelle Zugriffsstruktur ist, haben folgende Registerfunktion: Sie garantieren erstens die einzige alphabetische Zugriffsmöglichkeit auf alle Daten im Wörterverzeichnis und weisen zweitens — wie alle mediostrukturellen Register — eine Findfunktion auf. Genau die gleichen Registerfunktionen haben alphabetische zentrale Zugriffsregister in Wörterbüchern mit randomisierter Makrostruktur, wie z.B. Prosinger (1984).

In alphabetischen Wörterbüchern sind alle lemmatischen Angaben primär akzessiv; alle nichtlemmatischen Angaben sind sekundär akzessiv; auf die Letzteren kann daher nur zugegriffen werden, wenn vorher auf das Lemma zugegriffen wurde. Alphabetische zentrale Zugriffsregister in alphabetischen Wörterbüchern haben erstens die Funktion, dass auf die sekundär akzessiven Daten indirekt extern zugegriffen werden kann; ihre alphabetisch geordneten Registereingänge nennen daher nichtlemmatische Angaben, die zu einem oder mehreren Angabetypen je wörterbuchspezifischer Klassen von Angaben gehören. Welche Angabetypen das sind, ist meistens aus dem Registertitel ersichtlich. Es folgen einige Beispiele: Das zentrale alphabetische Zugriffsregister in Pfeifer (1993) hat den Titel: „Verzeichnis nicht an alphabetischer Stelle behandelte Wörter“. Ein Registerbeitrag lautet: „*abprotzen* Protze“. Ein Benutzer, der

anhand dieses Registereintrages einen Registerverweis erschließt und mit Hilfe des mediostrukturellen Leitelements |Protze| eine Verweispflichthandlung dadurch ausführt, dass er eine externe Registerzugriffshandlung ausführt, findet nach dem externen Zugriff auf das Lemma **Protze** im zugehörigen Wörterbuchartikel das nicht an alphabetischer Stelle lexikographisch bearbeitete sekundär akzessive Wort „a b p r o t z e n“, das am Schluss des Artikels wie folgt bearbeitet ist:

a b p r o t z e n Vb 'ein Geschütz von der Protze abheben und in Feuerstellung bringen` (Anfang 18. Jh.) 'ein Geschütz abfeuern` (18. Jh.).

Die wörterbuchspezifische Klasse von Angaben, zu der „a b p r o t z e n“ und alle mit den anderen Registereingängen genannten Angaben gehören, wird im Registertitel genannt; es ist die wörterbuchspezifische Klasse der „nicht an alphabetischer Stelle behandelten Wörter“. — Das zentrale alphabetische Zugriffsregister in Seibicke (1989) trägt den Titel „Register der in den Wörterbuchartikeln erwähnten gemeinsprachlichen und landschaftlichen Ausdrücke“. Aus diesem Register stammen die Registerausschnitte (1) und (2) in Abb. 1. Auch in diesem Register werden die sekundär akzessiven Angaben und dazu Teile von Angabetexten mit den Registereingängen genannt. Unter systematischen Aspekten sind die Register in Pfeifer (1993) und Seibicke (1989) gleichartig, und zwar auch hinsichtlich ihrer Registerfunktion. Beide weisen eine Findfunktion auf, die darin besteht, dass auf artikelinterne sekundär akzessive Daten, die zu einer wörterbuchspezifischen Klasse gehören, extern zugegriffen werden kann. Die Klassen können verschieden sein: Beispielsweise sind die Elemente der Klasse in Pfeifer (1993) wortbildungsmorphologisch zusammengehörige Wörter und in Seibicke (1989) sind es meistens Varietätenäquivalente. Der Terminus *Findfunktion* ist so zu verstehen, dass ein Register, das eine Findfunktion aufweist, dem Benutzer ermöglicht im Registerzugriffsbereich Daten zu finden und nicht etwa so, dass der Benutzer im Register etwas findet.

An dieser Stelle sei nun ausdrücklich gesagt, dass es gerade nicht die Aufgabe einer Registertheorie sein kann, alle inhaltlichen Möglichkeiten aufzulisten, die empirisch vorgefunden werden können. Vielmehr besteht die Aufgabe gerade darin, das strukturelle und funktionale Gemeinsame inhaltlich unterschiedlicher Fälle begrifflich zu erfassen. Für den Typ des zentralen alphabetischen Registers in alphabetischen Wörterbüchern mit Findfunktion für artikelintern sekundär akzessiven Daten, gibt es unzählige Ausprägungsmöglichkeiten, die sich nur dadurch unterscheiden, zu welcher Klasse die sekundär akzessiven Daten gehören, die mittels des Registers gefunden werden können. Bei der Planung von Registern kommt es daher darauf an, aus der großen Zahl dieser Möglichkeiten angemessen auszuwählen. Kriterien für eine begründete Auswahl sind erhältlich, wenn man überlegt, welche Register etwas dazu beitragen, die dem Wörterbuch zugeordneten Wörterbuchfunktionen (i.S.v. Wiegand 2001) besser zu erfüllen.

In alphabetischen Wörterbüchern treten weiterhin auch alphabetische

zentrale Register auf, die eine andere als die gerade behandelte Registerfunktion haben. Es handelt sich um die reduzierten Register, mit deren Registereingängen Lemmata genannt werden, wie z.B. in dem Register, aus dem (3) in Abb. 1 ein Registerausschnitt ist. In Sprachwörterbüchern sind Register dieses Typs relativ selten; im Sachwörterbüchern sind sie nach meinem Eindruck öfters anzutreffen. Beispielsweise findet man in Sportlexika reduzierte Personenregister, in denen die Namen der Sportler alphabetisch aufgelistet sind, die einen Personenartikel mit ihrem Namen als Lemma erhalten haben. Solche Register haben eine Registerfunktion, die man als Übersichtsfunktion zu einer Menge von ausgewählten Wörterbuchartikeln betrachten kann.

Im Folgenden betrachten wir alphabetische periphere Zugriffsregister am Beispiel eines umtextorientierten Zugriffsregisters. Der Registerausschnitt (10) in Abb. 2 stammt aus einem Register dieses Typs. Es ist das von mir erarbeitete „Register zur lexikographisch-historischen Einführung“ im Dornseiff (2004). Dieses Zugriffsregister funktioniert wie ein Sach- und Personenregister in einer fachlichen Monographie. Es dient also der Auffindung von Textstellen anhand von Termini und Namen und hat damit eine Finde- und Inhaltserschließungsfunktion.

Als nächstes werden die orientierenden Zugriffsregister kurz betrachtet. Während in nichtorientierenden Zugriffsregister auf die Registereingänge ausschließlich Registerverweisangaben folgen, so dass nichtorientierende Zugriffsregister nur eine Findefunktion aufweisen, treten in orientierenden Zugriffsregistern auch angereicherte Registerinträge auf, wie beispielsweise in dem Register, aus dem der Registerausschnitt (9) in Abb. 1 stammt. Orientierende Zugriffsregister haben daher nicht nur eine Findefunktion. Vielmehr erhält der Benutzer auf bestimmte Suchfragen eine inhaltliche Antwort allein anhand der Registerangaben. Aus diesem Grund heißt die Registerzugriffsstruktur von alphabetischen zentralen orientierenden Zugriffsregistern auch *alphabetische zentrale adressen- und inhaltsvermittelnde Registerzugriffsstruktur*. Die orientierenden Zugriffsregister haben mithin neben der Finde- eine Auskunftsfunktion. — Damit sind die wichtigsten der mediostrukturellen Register hinsichtlich ihrer Registerfunktionen behandelt, und wir werfen noch einen kurzen Blick auf die nichtmediostrukturellen Register.

Für alle nichtmediostrukturellen Register gilt — da sie keine Zugriffsregister sind — dass sie keine Findefunktion aufweisen. Sie haben nur einen direkten Zugriffsbereich, in dem der Benutzer Daten direkt auffinden kann. In einem rückläufigen Register, das nur Registereingänge aufweist, kann er viele Wörter finden, die auf *-at* enden. Daher gilt: Alle nichtmediostrukturellen Register weisen ausschließlich eine themengebundene Auskunftsfunktion auf. — Zusammenfassend kann also festgestellt werden: Alle mediostrukturellen Register weisen eine Findefunktion auf. Einige mediostrukturelle Register, die zu bestimmten Registertypen gehören, weisen zusätzlich eine Auskunftsfunktion auf. Alle nichtmediostrukturellen Register haben nur eine Auskunftsfunktion. Alle Register tragen etwas dazu bei, dass

die externe Datenakzessivität erweitert wird. Alle alphabetischen zentralen Zugriffsregister machen ein alphabetisches Printwörterbuch polyakzessiv.

4. Registerkritik

Das in diesem Beitrag bereitgestellte begriffliche Instrumentarium kann auch genutzt werden, um an vorhandenen Registern begründet hinsichtlich ihrer Anlage und hinsichtlich ihrer Funktionen Kritik zu üben und Verbesserungsvorschläge zu machen. Dies sei im Folgenden an einem Beispiel gezeigt.

Das „Lexikon zur Wortbildung der deutschen Sprache (Augmentation und Diminution)“ (Karbelaschwili 2001), das für Benutzer gedacht ist, die Deutsch als Fremdsprache lernen, ist ein Lehrstück dafür, wie ein Wörterbuchregister nicht angelegt sein sollte. Das Wörterbuch ist sachlich gegliedert und weist somit eine nichtformgeprägte Makrostruktur auf. Die sachliche Gliederung geht aus dem Ausschnitt des Inhaltsverzeichnisses hervor, der in Abb. 18 zu sehen ist.

(25)

1. Augmentation		1. Diminution	
1.1	Substantiv	22	2.1 Substantiv..... 140
1.1.1	Adposition	22	2.1.1 Adposition..... 140
1.1.2	Postposition	71	2.1.2 Postposition..... 151
1.2	Adjektiv	85	2.2 Adjektiv
1.2.1	Adposition	85	2.2.1 Adposition..... 154
1.2.2	Postposition	132	2.2.2 Postposition..... 158
1.3	Verb.....	134	2.3 Verb
1.3.1	Adposition	134	2.3.1 Adposition..... 159
			2.3.2 Postposition..... 161

Abb. 18: Ausschnitt aus dem lexikographischen Inhaltsverzeichnis von Karbelaschwili (2001)

Karbelaschwili (2001) weist eine gestaffelte numerischsachbezogene/alphabetische Zugriffsstruktur auf. Ein Benutzer, der ein Affix oder Affixoid nachschlagen möchte, kann das nur dann, wenn er in der Lage ist, das Affix oder Affixoid in das sachliche System einzuordnen, denn obwohl es sich um ein nichtalphabetisches Wörterbuch handelt, hat es kein alphabetisches Zugriffsregister. Angenommen, ein Benutzer möchte das augmentative Präfixoid *Bären-* nachschlagen, dann muss er es zunächst richtig klassifizieren, so dass er weiß, dass er *Bären-* unter 1.1.1 (vgl. Abb. 18) suchen muss. Das externe Zugreifen verläuft dann wie folgt: Erst muss der Benutzer im lexikographischen Inhaltsverzeichnis die Seitenzahl ermitteln, dann muss er auf die genannte Seite 22 zugreifen und danach auf den Sachgruppentitel. Innerhalb der Sachgruppe findet der Benutzer dann eine alphanumerische äußere Doppelzugriffsstruktur, die z.B. folgende äußeren Zugriffstextelemente aufweist:

<1. ASS- , 2. AB- , 3. ABER- , 4. ADLER- , 5. AFFEN- , 6. AFTER- , ... ,
21. BÄREN- , ... , 281 ZENTRAL- >.

Jede Sachgruppe hat eine eigene Zugriffsstruktur dieser Art. Die vorgeordnete numerische Zugriffsstruktur ist allerdings akzessiv funktionslos, denn sie wird in einem Zugriffsregister nicht genutzt; darin besteht aber gerade der Sinn einer alphanumerischen äußeren Doppelzugriffsstruktur, wie man in Ildikó (2005) sehen kann (vgl. Registerausschnitt (23) in Abb. 15). Statt eines Zugriffsregisters weist Karbelaschwili (2001) ein so genanntes (nichtmediostrukturelles) Elementenregister auf, das ein reduziertes Register ist und mithin nur aus Registereingängen besteht. In diesem sind die Sachgruppentitel, die Affixe und Affixoide genau so angeordnet, wie in der Makrostruktur! In dem Register kann nicht ohne weiteres nachgeschlagen werden. Vielmehr muss der Benutzer die gleiche klassifikatorische Vorarbeit leisten, die er erbringen muss, wenn er im Wörterverzeichnis nachschlagen möchte; erst danach kann er in elf aufeinander folgenden Teilen, die jeweils alphabetisch sind, systematisch suchen. Was er nach erfolgreicher Suche findet, ist dann allerdings nichts weiter als das Affix! Die Registerinformation, die erhältlich ist, besteht mithin nur darin, dass das Affix primär gebucht ist!

Richtig und akzessiv funktional wäre es gewesen, wenn man die Lemmata vollständig und damit sachgruppenübergreifend durchnummeriert und dann ein alphabetisches Zugriffsregister angelegt hätte, in dessen Registerinträgen die Lemmanummern mit einer numerischen Registerverweisangabe genannt werden. Die Registerinträge hätten dann folgende Form: Aas- 1 , Ab- 2 , aber- 3 , Adler- 4 , -affe 5, Affen- 5. Dies wären mithin einfache reduzierte Registerinträge. Ein alphabetisches zentrales Zugriffsregister dieser Art macht die klassifikatorische Vorarbeit des Benutzers überflüssig und erlaubt den schnellen punktgenaue registervermittelten Zugriff auf alle Affixe und Affixoide. Statt der einfachen reduzierten kann man auch erweiterte angereicherte Registerinträge folgender Form anlegen: „Aas- 1; Aug. Subst. Adpos.“ (mit „Aug.“ als Abkürzung für *Augmentation*; mit „Subst.“ als Abkürzung für *Substantiv* und mit „Adpos.“ als Abkürzung für *Adposition*). Dann hätte das Register nicht nur eine Findenfunktion, sondern auch eine eigenständige Auskunftsfunktion, da Antworten auf Fragen nach der Klassenzugehörigkeit erhältlich sind.

5. Schlussbemerkung

Nach meiner Kenntnis der empirischen Basis, die freilich in einem Beitrag dieser Art nur recht fragmentarisch präsentiert werden kann, reichen die hier dargestellten Grundlagen einer Theorie der Register wahrscheinlich aus, um jedes vorgelegte Wörterbuchregister lexikographietheoretisch erklären und einordnen zu können, das zu einem Printwörterbuch gehört, das in einer Sprache verfasst ist, die mittels einer Alphabetschrift so verschriftet ist, dass die zugehörigen Texte eine dextrograde Schriftrichtung aufweisen.

Literatur

- Augst, Gerhard.** 1984. *Kinderwort. Der aktive Kinderwortschatz (kurz vor der Einschulung) nach Sachgebieten geordnet mit einem alphabetischen Register.* Frankfurt a.M. [etc.]: Peter Lang.
- Augst, Gerhard.** 2003. *Dialektwörterbuch der Verbandsgemeinde Altenkirchen.* Wölmersen: Arbeitskreis für Heimatgeschichte und Brauchtumspflege e.V.
- Blair, Ann.** 2000. Annoting and Indexing Natural Philosophy. Frasca-Spada, M. und N. Jardine (Hrsg.). *Books and the Sciences in History:* 69-89. Cambridge: Cambridge University Press.
- Bratschi, Armin und Rudolf Trüb.** 1991. *Simmentaler Wortschatz. Wörterbuch der Mundart des Simmentals (Berner Oberland / Mit einer grammatischen Einleitung und mit Registern.* Unter Mitarbeit v. Lily Trüb sowie v. Maria Bratschi u. Ernst Max Perren. Zeichnungen v. Rolf Oberhäusli. Grammatiken und Wörterbücher des Schweizerdeutschen in allgemeinverständlicher Darstellung. Bd. XII. Thun: Ott Verlag.
- Brunner, Horst und Rainer Moritz. (Hrsg.).** 1997. *Literaturwissenschaftliches Lexikon.* Grundbegriffe der Germanistik. Berlin: Erich Schmidt.
- Bustos Plaza, Alberto und Herbert Ernst Wiegand.** 2006. Condensación textual lexicográfica: esbozo de una concepción integral. *Revista de Lexicografía* 12: 7-46.
- Dobel, Richard (Hrsg.).** 1972. *dtv-Lexikon der Goethe-Zitate.* Bd. 1: A-N; Bd. 2: O-Z. München: Deutscher Taschenbuchverlag.
- Dornseiff, Franz.** 2004. *Der deutsche Wortschatz nach Sachgruppen.* 8., völlig neu bearb. und mit einem vollständigen alphabetischen Zugriffsregister versehene Aufl. von Uwe Quasthoff. Mit einer lexikographisch-historischen Einführung und einer ausgewählten Bibliographie zur Lexikographie und Onomasiologie von Herbert Ernst Wiegand. Berlin/New York: Walter de Gruyter.
- Duden-DaF = Dudenredaktion (Hrsg.). 2002. *Standardwörterbuch Deutsch als Fremdsprache.* Mannheim [etc.]: Dudenverlag.
- Duden-⁴GFWB = Wiss. Rat der Dudenredaktion (Hrsg.). 2007. *Duden. Das große Fremdwörterbuch. Herkunft und Bedeutung der Fremdwörter.* 4., aktual. Aufl. Mannheim [etc.]: Bibliographisches Institut & F.A. Brockhaus AG.
- Duden-GW = Duden. 1976-1981. *Das große Wörterbuch der deutschen Sprache in sechs Bänden.* Hrsg. und bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter der Leitung von Günther Drosdowski. Mannheim [etc.]: Bibliographisches Institut.
- Duden-10 = Dudenredaktion (Hrsg.). 2002. *Duden. Das Bedeutungswörterbuch.* 3., neu bearb. u. erw. Aufl. Der Duden in zwölf Bänden. Bd. 10. Leipzig/Wien/Zürich: Dudenverlag.
- Erk, Heinrich.** 1972. *Zur Lexik wissenschaftlicher Fachtexte. Verben — Frequenz und Verwendungsweise.* Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts 4. München: Max Hueber.
- Eymer, Wilfrid.** 1997. *Eymers Pseudonymen Lexikon. Realnamen und Pseudonyme in der deutschen Literatur.* Teil 1: Realnamen mit den verwendeten Pseudonymen. Teil 2: Pseudonyme mit Verweisen auf die Realnamen. Bonn: Kirschbaum Verlag.
- Friederich, Wolf.** 1966. *Moderne deutsche Idiomatik. Systematisches Wörterbuch mit Definitionen und Beispielen.* München: Max Hueber.

- FWB-1 = Anderson, Robert R., Ulrich Goebel und Oskar Reichmann (Hrsg.). *Frühneuhochdeutsches Wörterbuch*. Bd. 1: *Einführung, a-äpfelkern*. Bearb. v. Oskar Reichmann. Berlin/New York: Walter de Gruyter.
- FWB 8/1 = Goebel, Ulrich und Oskar Reichmann (Hrsg.). 1997. *Frühneuhochdeutsches Wörterbuch*. Begr. v. R. R. Anderson, U. Goebel, O. Reichmann. Bd. 8. Lief. 1 i, j. Bearb. v. V. Winge. Berlin. New York: Walter de Gruyter.
- Götz, Dieter und Hans Wellmann (Hrsg.)**. 2007. *Langenscheidt. Taschenwörterbuch Deutsch als Fremdsprache. Das einsprachige Lernerwörterbuch für Einsteiger*. Neubearbeitung. In Zusammenarbeit mit der Langenscheidt-Redaktion. Berlin [etc.]: Langenscheidt.
- Hartmann, R.R.K. und Gregory James**. 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Heinzmann, Wilhelm**. 1993. *Wörterbuch der Pflanzennamen im Altwestnordischen*. Ergänzungsbände zum Reallexikon der Germanischen Altertumskunde. Bd. 7. Berlin. New York: Walter de Gruyter.
- Helbig, Gerhard und Wolfgang Schenkel**. 1980. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: VEB Bibliographisches Institut.
- Herwig, Axel**. 2005. *Das Kasseler Mundartwörterbuch*. Überarb. v. N. Rose, E. Reppel und E. Schmidt. Gudensberg-Gleichen: Wortberg Verlag.
- Herzog, Hans-Ulrich und Georg Hannes**. 1990. *Lexikon Flaggen und Wappen*. Leipzig: VEB Bibliographisches Institut.
- Ildikó, Fata**. 2005. *Ungarisch-Deutsches/Deutsch-Ungarisches Fachwörterbuch zur Rentenversicherung*. Szeged: Grimm Kiadó.
- Karbelaschwili, Samson**. 2001. *Lexikon zur Wortbildung der deutschen Sprache (Augmentation und Diminution)*. 2. korrig. u. erw. Aufl. Materialien Deutsch als Fremdsprache 51. Regensburg: Armin Wolff. Univ. Regensburg.
- Kirkness, Alan (Hrsg.)**. 1988. *Deutsches Fremdwörterbuch*. Begonnen v. Hans Schulz, fortgeführt v. Otto Basler, weitergeführt im Institut für deutsche Sprache. 7. Bd. Berlin/New York: Walter de Gruyter.
- Kluge, Friedrich**. 1943. *Etymologisches Wörterbuch der deutschen Sprache*. 12. u. 13. unveränd. Aufl. Mit Unterstützung durch W. Krause bearb. v. A. Götze. Berlin/Leipzig: Walter de Gruyter.
- Kluge, Friedrich und Elmar Seebold**. 1995. *Etymologisches Wörterbuch der deutschen Sprache*. 23. erw. Aufl. Berlin/New York: Walter de Gruyter.
- Kytzler, Bernhard und Lutz Redemund**. 2007. *Unser tägliches Latein. Lexikon des lateinischen Spracherbes*. 7. Aufl. Mainz: Philipp von Zabern.
- Meng, Heinrich**. 1986. *Mundartwörterbuch der Landschaft Baden im Aargau nach Sachgruppen*. Grammatiken und Wörterbücher des Schweizerdeutschen in allgemeinverständlicher Darstellung. Bd. X. Baden: Baden Verlag.
- Müller, Wolfgang**. 1973. *Duden. Leicht verwechselbare Wörter*. Mannheim/Wien/Zürich: Dudenverlag.
- NLO 1997 = Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt, Dag Gundersen, Jón Hilmar Jónsson und Bo Svensén. *Nordisk Leksikografisk Ordbog*. Skrifter utgitt av Nordisk forening for leksikografi 4. Oslo: Universitetsforlaget AS.

- Paul, Hermann.** 2002. *Deutsches Wörterbuch. Bedeutungsgeschichte und Aufbau unseres Wortschatzes.* 10. überarb. u. erw. Aufl. v. Helmut Henne, Heidrun Kämper und Georg Objartel. Tübingen: Max Niemeyer.
- Pfeffer, J. Alan.** 1987. *Deutsches Sprachgut im Wortschatz der Amerikaner und Engländer. Vergleichendes Lexikon mit analytischer Einführung und historischem Überblick.* Tübingen: Max Niemeyer.
- Pfeifer, Wolfgang.** 1993. *Etymologisches Wörterbuch des Deutschen.* 1. Bd. A–L; 2. Bd.: M–Z. 2. Aufl. durchges. und erg. v. Wolfgang Pfeifer. Berlin: Akademie Verlag.
- Prosinger, Wolfgang.** 1984. *Das rabenstarke Lexikon der Scene-Sprache. Der große Durchblick für alle Freaks, Spontis, Schlaffis, Softies, Flipper und Hänger sowie deren Verwandte und sonstige Fuzzis.* Illustriert von P. Gaymann. Frankfurt am Main: Eichborn Verlag.
- Riehme, Joachim.** 1990. *Gleich gesprochen — verschieden geschrieben. Zum Verwecheln ähnliche Wörter und ihre richtige Schreibung.* Leipzig: VEB Bibliographisches Institut.
- Rouse, Mary und Richard H. Rouse.** 1982. *La naissance des index.* Martin, H.-J. und R. Chartier (Hrsg.). *Histoire d'édition française: 77-85.* Paris: Promodes.
- Scheuplein-Fritz, M., A. König, S. Krämer-Neubert und N.R. Wolf (Hrsg.).** 1996. *Wörterbuch von Unterfranken. Eine lexikographische Bestandsaufnahme.* Würzburg: Königshausen & Neumann.
- Schierholz, Stefan und Herbert Ernst Wiegand.** 2005. *Die Wörterbücher zur Sprach- und Kommunikationswissenschaft. Eine neue Konzeption der linguistischen Fachlexikographie und ihre computergestützte Praxis.* *Lexicographica* 20: 164-264.
- Schröder, Jochen.** 1992. *Lexikon deutscher Präfixverben.* Berlin [etc.]: Langenscheidt. Verlag Enzyklopädie.
- Seibicke, Wilfried.** 1989. *Duden. Wie sagt man anderswo? Landschaftliche Unterschiede im deutschen Sprachgebrauch.* 2. neu bearb. u. erw. Aufl. Duden-Taschenbücher. Bd. 15. Mannheim/Wien/Zürich: Dudenverlag.
- WDG-1 = Klappenbach, Ruth und Wolfgang Steinitz (Hrsg.). 1974. *Wörterbuch der deutschen Gegenwartssprache.* 1. Bd: A–deutsch. Bearb.: R. Klappenbach u. H. Malige-Klappenbach. 7. Aufl. Berlin: Akademie-Verlag.
- Wiegand, Herbert Ernst.** 1983. *Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung.* Wiegand, H.E. (Hrsg.). *Studien zur neuhochdeutschen Lexikographie III:* 401-474. *Germanistische Linguistik* 1–4/82. Hildesheim/Zürich/New York: Georg Olms Verlag.
- Wiegand, Herbert Ernst.** 1995. *Lexikographische Texte in einsprachigen Lernerwörterbüchern. Kritische Überlegungen anlässlich des Erscheinens von Langenscheidts Großwörterbuch Deutsch als Fremdsprache.* Popp, Heidrun (Hrsg.). *Deutsch als Fremdsprache. An den Quellen eines Faches. Festschrift für Gerhard Helbig zum 65. Geburtstag:* 463-499. München: Judicium Verlag.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* 1. Teilband. Mit 159 Illustrationen im Text. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998a. *Altes und Neues zur Makrostruktur in alphabetischen Printwörterbüchern.* Wiegand, H.E. (Hrsg.). *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium:* 348-372. *Lexicographica. Series Maior* 84. Tübingen: Max Niemeyer.

- Wiegand, Herbert Ernst.** 1998b. Lexikographische Textverdichtung. Entwurf einer vollständigen Konzeption. Zettersten, Arne, Viggo Hjørnager Pedersen und Jens Erik Mogensen (Hrsg.). 1998. *Symposium on Lexicography VIII. Proceedings of the Eighth International Symposium on Lexicography, May 2–4, 1996, at the University of Copenhagen*: 1-35. *Lexicographica*. Series Maior 90. Tübingen: Max Niemeyer.
- Wiegand, Herbert Ernst.** 2000. Adressierung in der ein- und zweisprachigen Lexikographie. Eine einführende Übersicht über die Forschungs- und Problemlage. *Lexikos* 10: 32-74.
- Wiegand, Herbert Ernst.** 2000a. Über tabellarische Wörterverzeichnisse und deren Wörterbuchartikel. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 16: 212-234.
- Wiegand, Herbert Ernst.** 2001. Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neusten Wörterbuchforschung. *Lexicographica* 17: 217-248.
- Wiegand, Herbert Ernst.** 2002. Altes und Neues zur Mediostruktur in Printwörterbüchern. *Lexicographica* 18: 168-252.
- Wiegand, Herbert Ernst.** 2002a. Adressierung in zweisprachigen Printwörterbüchern. Wiegand, H.E. (Hrsg.). 2002. *Studien zur zweisprachigen Lexikographie mit Deutsch VIII*: 111-175. *Germanistische Linguistik* 166. Hildesheim/Zürich/New York: Georg Olms 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.** 2003a. Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research. Städtler, Th. (Hrsg.). 2003. *Wissenschaftliche Lexikographie im deutschsprachigen Raum. Im Auftrag der Heidelberg Akademie der Wissenschaften*: 417-437. Heidelberg: Winter Verlag.
- Wiegand, Herbert Ernst.** 2005. Über die Datenakzessivität in Printwörterbüchern. Einblicke in neuere Entwicklungen einer Theorie der Wörterbuchform. *Lexikos* 15: 196-230.
- Wiegand, Herbert Ernst.** 2005a. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 21: 202-379.
- Wiegand, Herbert Ernst.** 2006. Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörtertyp am Beispiel der *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)*. *Lexikos* 16: 205-221.
- Wiegand, Herbert Ernst.** 2006a. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 22: 187-261.
- Wiegand, Herbert Ernst.** 2007. Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform. *Lexikos* 17: 180-211.
- Wiegand, Herbert Ernst.** 2008. Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. *Lexicographica* 24. [erscheint.]
- Zedelmaier, Helmut.** 2004. „Facilitas inveniendi“. Zur Pragmatik alphabetischer Buchregister. Stammen, Th. und W.E.J. Weber (Hrsg.). *Wissenssicherung, Wissensordnung und Wissensverarbeitung. Das europäische Modell der Enzyklopädien*: 191-203. *Colloquia Augustana*. Bd. 18. Berlin: Akademie Verlag.
- Zehetner, Ludwig.** 1997. *Bairisches Deutsch. Lexikon der deutschen Sprache in Altbayern*. München: Heinrich Hugendubel.

Improving the Computational Morphological Analysis of a Swahili Corpus for Lexicographic Purposes*

Guy De Pauw, *CNTS — Language Technology Group, University of Antwerp, Antwerp, Belgium; School of Computing and Informatics, University of Nairobi, Nairobi, Kenya; and Xhosa Department, University of the Western Cape, Bellville, Republic of South Africa (guy.depauw@ua.ac.be)*,

and

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

Abstract: Computational morphological analysis is an important first step in the automatic treatment of natural language and a useful lexicographic tool. This article describes a corpus-based approach to the morphological analysis of Swahili. We particularly focus our discussion on its ability to retrieve lemmas for word forms and evaluate it as a tool for corpus-based dictionary compilation.

Keywords: LEXICOGRAPHY, MORPHOLOGY, CORPUS ANNOTATION, LEMMATIZATION, MACHINE LEARNING, SWAHILI (KISWAHILI)

Samenvatting: **Accuratere computationele morfologische analyse van een Swahili corpus voor lexicografische doeleinden.** Computationale morfologische analyse is een belangrijke eerste stap in de automatische verwerking van natuurlijke taal en een nuttig lexicografisch hulpmiddel. Dit artikel beschrijft een corpusgebaseerde aanpak voor de morfologische analyse van het Swahili. We concentreren ons hierbij vooral op de lemmatiseringseigenschappen van het ontwikkelde systeem en evalueren het als een hulpmiddel bij de corpusgebaseerde ontwikkeling van woordenboeken.

Sleutelwoorden: LEXICOGRAFIE, MORFOLOGIE, CORPUSANNOTATIE, LEMMATISERING, AUTOMATISCHE LEERTECHNIEKEN, SWAHILI (KISWAHILI)

* An earlier version of this article was presented at the Thirteenth International Conference of the African Association for Lexicography, organized by the Bureau of the *Woordeboek van die Afrikaanse Taal*, Stellenbosch, Republic of South Africa, 1–3 July 2008.

1. Bantu computational lexicography

The last couple of years have seen a definite empirical shift in Bantu lexicography. The integration of corpus data in the arduous process of dictionary compilation allows the lexicographer to semi-automatically unearth examples for the dictionary entries in actual language use. It has become unimaginable to compile a wide-coverage dictionary for a Bantu language without the use of a large language corpus and a functional corpus query package (CQP). In De Schryver and De Pauw (2007) it was shown how the fields of natural language processing (NLP) and lexicography can collaborate towards enhancing the functionality of a CQP, by integrating a fast and accurate data-driven part-of-speech (POS) tagger.

In this article, we investigate how another typical NLP component — namely morphological analysis — can be developed with a minimal amount of manual effort, and demonstrate how it can be used as a CQP component. As a case study, we choose Swahili, a widely spoken Bantu language with no (publicly accessible) morphological analyzer or morphologically annotated lexicon. We will show how both of these resources can easily be developed using a machine-learning approach.

In Section 2, we describe some of the current approaches to morphological analysis and provide a comprehensive overview of previous work on Bantu languages. We then discuss, in Section 3, the construction of a Swahili morphological database, which will be used as an information source for the machine-learning approach described in Section 4. After a quantitative evaluation of the system, in Section 5, we conclude, in Section 6, with a discussion of the current state of affairs and some pointers to future work.

2. Computational morphological analysis

Computational morphological analysis is an important first step in the automatic treatment of natural language. Finding the *minimal meaning bearing units* that constitute a word, can provide a wealth of linguistic information that becomes useful when processing the text on other levels of linguistic description, such as phonology, syntax and even semantics.

In most practical language technology applications, morphological analysis is used to perform *lemmatization*. A typical application of a lemmatizer is integrated in *Google's* search facility, which automatically lemmatizes a search term like 'discussions' to also produce hits for the word form 'discussion'. Lemmatization is also often used to enhance statistical models of language in other language technology applications, like machine translation (Oflazer 2008) and speech recognition (De Pauw et al. 2004). There are however few publications that explicitly discuss the obvious lexicographic application of a lemmatizer, i.e. as a CQP component (but see Christ 1994, Kilgarriff et al. 2008). In this article, we will therefore focus our discussion on the lemmatization abilities of

the developed system and evaluate it as a tool for corpus-based dictionary compilation.

2.1 Current approaches

The most widely used approach to computational morphology uses the *two-level formalism* (Koskenniemi 1983). This rule-based method typically operates on the character level and associates each character with a given morphological property. The approach distinguishes between the surface and lexical realizations of a given morpheme (hence two-level) and attempts to establish a mapping between the two. The two-level formalism uses a (large) collection of finite-state transducers which each implement a particular morphological rule. While the framework itself is language independent, these rules typically need to be manually constructed for each language and/or sub-domain of a language, which makes the development of such a morphological analyzer very costly and time-consuming.

In the late nineties a few interesting corpus-based alternatives have surfaced. Rather than requiring expert linguistic knowledge to construct a morphological analyzer, these approaches automatically induce the required information from a morphologically annotated data set, such as CELEX (Baayen et al. 1995). Using statistical processing (Masaaki 1999) and/or machine-learning techniques (Van den Bosch and Daelemans 1999), these data-driven methods establish an effective and truly language-independent technique for morphological analysis, that can easily be ported to new domains. Furthermore, manually constructed rule-based analyzers typically do not significantly outperform data-driven approaches in a direct comparison (De Pauw et al. 2004).

In more recent years, research on computational morphology has mainly concentrated on unsupervised approaches. These methods attempt to automatically induce the morphological properties of a language on the basis of raw, unannotated text, using minimum-distance edit metrics and pattern-matching techniques.

2.2 Bantu computational morphological analysis

While great advances have been made for many Indo-European and Asian languages, most computational morphological models for Bantu languages are still in the developmental stage. This is not only due to the relatively limited commercial interest in these languages, but also because of the often intricate morphology, which renders both the construction of rule-based and data-driven methods troublesome.

Most of the research on computational morphology of Bantu languages is being conducted in South Africa and is rooted in the rule-based two-level formalism. Morphological analyzers are being developed for Northern Sotho

(Kotzé and Anderson 2005, Bosch et al. 2006, Anderson et al. 2007), Zulu (Tajard and Bosch 2005, Bosch et al. 2006, Pretorius and Bosch 2007), Xhosa, Swazi and Tswana (Bosch et al. 2006). Smaller projects have also looked into aspects of the morphology of Shona (Ridings and Mavhu 2002), Zimbabwean Ndebele (Maphosa 2002), Kwanyama (Hurskainen and Halme 2001) and Rwanda (Muhirwe 2007).

The rule-based two-level morphology formalism has also been applied to the verbal morphology of Gusii (Elwell 2006) and to Swahili (Hurskainen 1992, 1996, 2004). The latter rule-based morphological analyzer for Swahili is known as SALAMA, and was used to lemmatize the *Helsinki Corpus of Swahili* (HCS, Hurskainen 2004a). This system is not publicly available, however.

The user interface for two Swahili dictionaries on the Internet, viz. the *Kamusi Project* and the *Online Swahili-English Dictionary* (Hillewaert, Joffe and De Schryver 2008), integrate rule-based morphological analyzers. These are useful for dictionary queries, but are limited to analyzing (verb) forms for which the underlying lemma is also present in the dictionary.

Data-driven approaches are indeed few and far between, with some notable exceptions. A data-driven morpho-syntactic tagger was developed for Swahili¹ (De Pauw et al. 2006) and Northern Sotho² (De Schryver and De Pauw 2007). An unsupervised approach to morphological analysis has been applied to Luo, a Nilotic language (De Pauw et al. 2007) and Gikuyu (De Pauw and Wagacha 2007). The latter actually constitutes a viable alternative to unsupervised methods such as *AutoMorphology* (Goldsmith 2001) or *Morfessor* (Creutz et al. 2005), which are not well equipped to handle Bantu morphology. Compare also with Elwell (2008).

Lindén (2008) describes a semi-supervised method for the lemmatization of Swahili words. The method uses the annotation of HCS to induce a probabilistic model that is able to guess base forms of previously unseen words.

Finally, Elwell (2008) describes a novel technique for verbal morphological analysis of Swahili. It uses the insight that Swahili morphemes are open syllables and monosyllabic to create a maximum entropy-based classifier that categorizes syllables for different aspects of the verbal morphology. While limited to verbal morphology only, it is to our knowledge the only machine-learning approach to morphological analysis that specifically caters to a Bantu language in terms of knowledge representation.

3. Towards a Swahili morphological database

The research described in this article wants to fill the void by creating a data-driven morphological analyzer for Swahili that handles all morphologically productive word classes. To this end, one needs a morphologically annotated word list. While this is not available as such, one can go a long way by extracting the necessary information from HCS, lemmatized using the SALAMA morphological analyzer.

In HCS, every word is associated with its lemma, POS-tag, some morphological features and an English translation, like in examples (1) and (2).

- | | | | | | | | | |
|-----|-------------------|---------|---|-------------|-------|------|-----------|--------------------------------|
| (1) | ulikanusha | kanusha | V | [1/2-SG2-SP | VFIN | PAST | SV | EXT: SVO-C |
| | | | | CAUS:sh | :EXT] | | | deny, disprove, refute, negate |
| (2) | ulikoanzia | anza | V | [1/2-SG2-SP | VFIN | PAST | 15-SG-REL | SV SVO |
| | | | | EXT: APPL | :EXT] | | | begin, establish |

We can use this information to perform pattern-matching and match the lemma to the word form. Through this operation we can automatically induce a morphologically segmented surface and lexical representation of the word form, in which we distinguish a prefix group (**[P]**), the root morpheme (**[R]**) and a suffix group (**[S]**). In some cases, this is straightforward, as for the entry in example (1) which can easily be transformed into example (3).

- | | | | | |
|-----|-------------------|---------|---|--|
| (3) | ulikanusha | kanusha | → | Surface: uli [P] + kanusha [R] |
| | | | → | Lexical: uli [P] + kanusha [R] |

For the entry in example (2), this leads to the creation of a bound root morpheme *anz-* in the surface representation, associated with the full lemma *anza* in the lexical representation.

- | | | | | |
|-----|-------------------|------|---|---|
| (4) | ulikoanzia | anza | → | Surface: uliko [P] + anz [R] + ia [S] |
| | | | → | Lexical: uliko [P] + anza [R] + ia [S] |

Using this method, we automatically extracted a morphological database of 97 000 entries from the 9.7-million-word HCS. We retained word forms from morphologically productive word classes only, and filtered out noise as much as possible by discarding low-frequency tokens and English words. However, some misspelt words (e.g. *uuondoe*) and non-English loan words (e.g. *Deutsche*) still make up for some noise in the data that cannot be automatically discarded.

Since HCS has been lemmatized using an automated method, quite a few erroneous and inconsistent lemmatizations can be observed in the data. We therefore randomly extracted 10% of the data from the morphological database and had it manually annotated according to the prefix-root-suffix ([P]-[R]-[S]) protocol illustrated in examples (3) and (4). The availability of this manually annotated *gold-standard evaluation set* does not only allow us to cross-check the accuracy of our system on clean data, but also enables a *post-hoc* evaluation of the rule-based approach used to annotate HCS.

Similarly to the annotation approach described in De Schryver and De Pauw (2007), we used Microsoft Excel as the annotation environment. The annotation sheet seen in Figure 1 lists each word on a separate row. The word form itself is listed in Column A. Column B contains a sentence extracted from HCS, illustrating that word form in context. The minimized sentence can be displayed in full by double-clicking on the cell. Columns C and onwards list the individual characters of the word form from Column A, separated by blank cells.

Each blank cell has a drop-down box available with three options: **P** (end of prefix group), **R** (end of root group) and **S** (end of suffix group). The annotator can quickly move through the annotation process using only the keyboard or mouse clicks. In practice, the **S** annotation does not need to be indicated, as any character to the right of the root group automatically constitutes the suffix group (see e.g. line 7304). Furthermore, if the word does not have a suffix group, only the **P** annotation needs to be identified by the annotator, since what remains is automatically considered to be the root group (see e.g. line 7314).

In this way, the surface representation of the morpheme boundaries is annotated. In a second annotation step, the lexical representations of the roots, thus the actual lemmas, are double-checked and corrected where necessary.

	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	AAABACAD
7304	wakaleta		w	a	k	a	P	l	e	t	R	a															
7305	wakalima		w	a	k	a	P	l	i	m	R	a															
7306	wakamatwa		w	a	P	k	a	m	a	t	R	w	a														
7307	wakamilifu		w	a	P	k	a	m	i	l	i	R	f	u													
7308	wakamkuta		w	a	k	a	m	P	k	u	t	R	a														
7309	wakampigia		w	a	k	a	m	P	p	i	g	R	i	a													
7310	wakamtegemea		w	a	k	a	m	P	t	e	g	e	m	e	R	a											
7311	wakamuuliza		w	a	k	a	m	u	P	u	l	i	z	R	a												
7312	wakamwomba		w	a	k	a	m	w	P	o	m	b	R	a													
7313	wakandamizwaji		w	a	P	l	a	n	d	a	m	i	z	R	w	a	j	i									
7314	wakanishauri		w	a	P	l	a	n	i	P	s	h	a	u	r	i											
7315	wakanywa		w	a	P	l	a	n	y	R	w	a															
7316	wakaongea		w	a	k	a	P	o	n	g	e	R	a														

Figure 1: Excel sheet containing the material annotated by the annotator

4. The 'Memory-Based Swahili Morphological Analyzer' (MBSMA)

In this section, we describe our data-driven method for morphological analysis of Swahili, which is based on supervised machine learning. It reuses and refines the basic methodology coined in Van den Bosch and Daelemans (1999) which has been successfully applied to morphologically rich(er) languages such as Dutch (De Pauw et al. 2004) and Arabic (Van den Bosch et al. 2007). We use the data set described in Section 3 as our primary information source, and describe two systems.

4.1 Character-based morphological analysis

The first system directly ports the character-based approach of the original method (Van den Bosch and Daelemans 1999) to Swahili. The technique is

based on the machine-learning method of memory-based learning, which takes a database of *instances* as training material. These instances have to be represented as a fixed-length string of features, which describe the linguistic context of the token to be classified. Each instance is associated with a class, in this case a morphological category. The memory-based learning algorithm then stores this data in memory and classifies new, unseen instances, by comparing them to the ones in memory and extrapolating the class of the closest matching instance in memory.

For morphological processing, we extracted instances from the morphological database described in Section 3 as follows: for each character in each word form, we created a single instance that describes that character in its context. In the case of example (4) (uliko[**P**] + anz[**R**] + ia[**S**]), we can extract the ten instances displayed in Table 1. Each character is used once as a focus character (**F**) and associated with the five characters to its left (**L1**→**L5**) and the five characters to its right (**R1**→**R5**).

The *o* character in Instance 5, for example, is preceded by a dash (-) marking the word boundary and the characters *u*, *l*, *i* and *k*. It is followed by the characters *a*, *n*, *z*, *i*, *a*. This instance is then associated with a morphological classification, in this case **P**(refix), marking the fact that *o* is the last character of the prefix group. Similarly, the character *z* (Instance 8) is associated with the **R**(oot) class, and the *a* in Instance 10 with the **S**(uffix) class. Characters that do not mark the end of a morpheme are classed with the default category **0**.

The window size for the surrounding context is a parameter that needs to be optimized. Contexts that are too small or too large will hamper the performance of the classifier. The optimal window size of five characters before and after the focus characters was automatically established on the basis of comparative experiments on a development set.

Table 1: Character-based instances extracted from the morphological database

	L5	L4	L3	L2	L1	F	R1	R2	R3	R4	R5	CLASS
1	-	-	-	-	-	u	l	i	k	o	a	0
2	-	-	-	-	u	l	i	k	o	a	n	0
3	-	-	-	u	l	i	k	o	a	n	z	0
4	-	-	u	l	i	k	o	a	n	z	i	0
5	-	u	l	i	k	o	a	n	z	i	a	P
6	u	l	i	k	o	a	n	z	i	a	-	0
7	l	i	k	o	a	n	z	i	a	-	-	0
8	i	k	o	a	n	z	i	a	-	-	-	R+a
9	k	o	a	n	z	i	a	-	-	-	-	0
10	o	a	n	z	i	a	-	-	-	-	-	S

Furthermore, characters marked with an **R** classification can have an extra instruction, like in Instance 8 in Table 1, where the full class is **R+a**. The added **+a** instruction functions as an indication that the full lexical representation for this root morpheme needs to be *repaired* from the surface representation to the lexical representation by adding an *-a* to the end.

During actual morphological analysis, i.e. the morphological segmentation of previously unseen word forms, the words are similarly deconstructed and represented as instances using the same information. If, for example, we are to morphologically segment the previously unseen word *kulikoamuriwa* (kuliko[**P**] + amuriwa[**R**]), we classify the instances for each character. During the processing of this word, we will encounter the instance in Table 2, for which the morphological class is unknown.

Table 2: Instance to be classified

L5	L4	L3	L2	L1	F	R1	R2	R3	R4	R5	CLASS
k	u	l	i	k	o	a	m	u	r	i	??

This instance is compared to each and every instance in the training set, recorded by the memory-based learner. In doing so, the classifier will try to find that training instance in memory that most closely resembles it. For the instance in Table 2, this might be Instance 5 in Table 1, as they share six features (**L4**, **L3**, **L2**, **L1**, **F** and **R1**). The memory-based learner then extrapolates the **P** class of this training instance and *predicts* it to be the class of the new instance. Finally, in a post-processing phase, the words are recompiled and the predicted classes, i.e. morpheme boundaries, are inserted.

4.2 Syllable-based morphological analysis

The second version of the memory-based morphological analyzer moves away from the default level of the character and instead describes the problem on the level of the syllable. For Swahili, this has already been shown to constitute a relevant level of description (Elwell 2008).

Using syllables rather than characters as features does involve an extra pre-processing step, namely syllabification. We adopted the syllabification approach described in Ngugi, Okelo-Odongo and Wagacha (2005) and marked syllable boundaries for the words in our morphological database. This process is very precise for Swahili word forms, although mistakes are made on inflected loan words. Example (5) illustrates the syllabification process for the word *ulikoanzia*. Note that the syllable *zi* is never considered as a syllable within the representation, as it is split by a morpheme boundary that yields the desired bound root morpheme.

$$(5) \text{ uliko}[P] + \text{anz}[R] + \text{ia}[S] \rightarrow \text{u|li|ko}[P] + \text{a|n|z}[R] + \text{i|a}[S]$$

The rest of the processing remains the same. Instances are extracted in much the same way, except that the features now refer to syllables instead of single characters (see Table 3). Working on the syllable level also means fewer instances are being extracted, which helps speed up training times for the memory-based learner.

Table 3: Syllable-based instances extracted from the syllabified morphological database

	L5	L4	L3	L2	L1	F	R1	R2	R3	R4	R5	CLASS
1	-	-	-	-	-	u	li	ko	a	nz	i	0
2	-	-	-	-	u	li	ko	a	nz	i	a	0
3	-	-	-	u	li	ko	a	nz	i	a	-	P
4	-	-	u	li	ko	a	nz	i	a	-	-	0
5	-	u	li	ko	a	nz	i	a	-	-	-	R+a
6	u	li	ko	a	nz	i	a	-	-	-	-	0
7	li	ko	a	nz	i	a	-	-	-	-	-	S

5. Experiments and evaluation

In this section, we evaluate MBSMA. First, we look at its performance as an NLP tool *per se*, observing its accuracy as a morphological segmenter and lemmatizer. Next, we take a more qualitative look at the approach as a lexicographic tool.

5.1 Evaluation as an NLP tool

We are most interested in the accuracy of the morphological analyzer on previously unseen words: how well is the system able to morphologically segment and lemmatize unknown word forms? To investigate this, we perform *blind testing*, which involves partitioning the data in two parts: a 90% partition to train the system, and a 10% partition to evaluate it. For the latter, we use the manually annotated gold standard evaluation set, described in Section 3.

There are many experimental parameters to consider while building the system. The optimal combination of information source and algorithmic parameters can be established through thorough experimentation on the training set. At no point during this optimization process however, do we gauge the performance of the system on the evaluation set. This would not only produce artificially inflated accuracy scores in the final evaluation, but would also serve to *overfit* the system on one particular set of words.

We compare the accuracy of four different approaches to morphological segmentation and lemmatization of Swahili:

- **Morfessor** (Creutz et al. 2005): an unsupervised approach that takes a list of words (without annotation) and automatically induces a morphological model. This model is then used to segment the words in the evaluation set.
- **SALAMA^x** (Hurskainen 2004): the morphological analyzer used to lemmatize HCS. Since the SALAMA morphological analyzer itself is not publicly available, we reverse engineered its accuracy score by comparing the original annotation of the HCS annotations to the manually corrected annotation of the gold-standard evaluation set.

- **MBSMA-c**: the memory-based morphological analyzer working on the character level.
- **MBSMA-s**: the memory-based morphological analyzer working on the syllable level.

We will follow the standard approach of using word-error rate (WER) as our primary evaluation metric. It expresses the accuracy on the word-level, i.e. how many words have *not* been completely correctly segmented and lemmatized. In other words: the lower the WER, the better the system.

Table 4 displays the experimental results. As expected, **Morfessor** does not yield a great accuracy score. It is only able to completely correctly segment the surface representation of the words in the evaluation set 29.3% of the time. Morfessor is further hindered in its lemmatization accuracy, as it is unable to map surface representations onto lexical representations. It should be pointed out, however, that this is by far the most cost-effective system to develop, since it does not require any prior knowledge of the morphology of the language in question and thereby completely factors out the human element.

Table 4: Accuracy scores for Morfessor, SALAMA^x, MBSMA-c and MBSMA-s on the manually annotated evaluation set

	Segmentation of the surface representation	Further lemmatization
	WER	WER
Morfessor	70.7 %	73.6 %
SALAMA^x	11.7 %	12.0 %
MBSMA-c	13.3 %	13.6 %
MBSMA-s	11.6 %	11.7 %

We evaluated **SALAMA^x** by comparing the original HCS annotation to the manually annotated evaluation set. SALAMA^x obviously performs much better with a WER of 11.7%, an enormous error reduction over Morfessor. Most errors are made on the selection of the wrong lemma for a given word form, which further percolates into the segmentation.

The result for **MBSMA-c** shows that simply porting the original methodology described in Van den Bosch and Daelemans (1999) provides a functional data-driven morphological analyzer. However, this character-based approach is still significantly being outperformed by SALAMA^x.

Finally, when we move the level of description up to the syllable, **MBSMA-s**, the memory-based approach can be observed to slightly outperform SALAMA^x, establishing a small, but statistically significant reduction in WER on surface-level segmentation and a more substantial reduction for lemmatization.

This result may be surprising: how can a data-driven approach outperform the system that was used to create its information source? The answer to

this question lies in the generalization capabilities of the machine-learning technique. As previously mentioned and as further illustrated by the SALA-MA^x results in Table 4, quite a few erroneous analyses can be found in the annotation of HCS. Rather than completely mimicking the properties of the data the machine-learning approach uses to train its model, it implicitly generalizes over the data and filters out the noise.

A closer look at the output of MBSMA-s reveals some general tendencies. About half of the mistakes are made by MBSMA-s either by introducing a prefix group where there should not be one (36%) or by misjudging the length of the prefix group (18%). MBSMA-s fails to identify a prefix group 22% of the time and simply attaches it to the root. This also happens 18% of the time for the suffix group. Only very rarely is MBSMA-s unable to retrieve the right lemma from a correctly segmented surface representation, underlining its ability to restore a surface root form into its underlying lexical representation.

5.2 Evaluation as a lexicographic tool

Now that we have established the accuracy of the Swahili morphological analyzer on a purely quantitative basis, we turn to evaluating it as a lexicographic tool. In this discussion, we will focus on the lemmatization capabilities of the morphological analyzer and discuss processing times, retrievability and lemma discovery in the context of lexicography. We refer to two aspects of the lemmatizer as a lexicographic tool:

- The lemmatizer as a component in a CQP: we want the tool to be able to quickly and accurately lemmatize all the words in a corpus, so that example contexts for a given dictionary entry/lemma can easily be looked up and included in the description. We consider this to be an off-line task.
- The lemmatizer as a tool for digital dictionary consultation: a user should be able to input an inflected word form in the lookup interface. This word form is then lemmatized on the fly and its associated lemma is looked up in the dictionary. This is considered to be an online task.

Particularly the latter purpose requires the lemmatizer to be fast. **Processing times** are luckily quite favourable for MBSMA-s. On a standard Duo Core 2Ghz machine, the system is able to lemmatize over 70 words per second, using about 64Mb of internal memory. Speed and memory usage can be further optimized by using more efficient algorithmic parameters with only a minimal negative impact on its accuracy.

Apart from processing speed, we also identify another parameter to evaluate the lemmatizer as a lexicographic tool, namely **retrievability**, literally defined as its *ability* to *retrieve* the word forms for a given lemma. During dictionary compilation, we want to be able to provide the lexicographer with as many proper example sentences as possible for a given dictionary entry. The lemmatizer can be of great assistance in this task.

We can quantify the retrievability performance of a lemmatizer by running a controlled experiment: we group the word forms in the manually annotated evaluation set according to lemma. We consider only those lemmas that are associated with at least two distinct word forms in the evaluation set, like in example (6).

(6) **umba**: alituumba aliumba aliwaumba aliyemuumba

We then look at the output of the MBSMA-s lemmatizer and similarly group the word forms according to their predicted lemma, like in example (7).

(7) **umba**: aliumba aliwaumba aliyemuumba itayumba uliyumba

We notice three problems in (7): the word form *alituumba* which should have been in this list, is not there, while the word forms *itayumba* and *uliyumba* are erroneously associated with lemma *umba*. We can quantify the retrievability performance of the lemmatizer by calculating precision and recall:

- **Precision** counts the number of correct word forms retrieved by the lemmatizer and divides it by the total number of found lemmas. In example (7), precision would be $3/5$ or 60%.
- **Recall** again counts the number of word forms correctly associated by the lemmatizer to the lemma in question and divides it by the number of lemmas that should have been retrieved, i.e. the number of word forms in example (6). For the predictions in example (7) the lemmatizer obtains a recall score of $3/4$ or 75%.

If we perform this calculation for each lemma of the evaluation set and average the scores, we get some insight into the performance of the lemmatizer in terms of retrievability (see Table 5), and find that **MBSMA-s** compares favourably to **SALAMA^x**. The precision score expresses that nine out of ten word forms provided by the lemmatizer are proper inflections of the lemma, while the recall score shows that the lemmatizer on average fails to retrieve only two out of ten word forms for a given lemma.

Table 5: Quantification of retrievability of the lemmatizer

	Precision	Recall
SALAMA^x	89.4%	83.6%
MBSMA-s	92.4%	83.4%

A final aspect of the lemmatizer as a CQP tool relates to the **discovery** of new lemmas. Not only do we want the lemmatizer to relate word forms to existing lemmas, we also want it to be able to discover new lemmas not yet described in the dictionary.

To evaluate the lemmatizer from this perspective, we run another controlled experiment, again using the manually annotated evaluation set. We list all the lemmas in the evaluation set and subsequently remove all word forms associated with these lemmas in the training set. This means that the evaluation set solely consists of word forms for lemmas for which no linguistic evidence exists in the training set. After retraining the system on the new, disadvantaged training set, we can estimate the performance of MBSMA-s as a lemma discovery technique, by calculating how many of the lemmas in the evaluation set are still correctly found.

Table 6 displays the results for this experiment. We also provide extrapolated results from a similar experiment, using a probabilistic semi-supervised method (Lindén 2008). Note that this is not a direct comparison — since different evaluation techniques and data sets were used — and therefore only serves as a guideline to interpret the scores for MBSMA-s. The score expresses the percentage of *unknown* lemmas that have been correctly identified. The results for MBSMA-s are encouraging.

Table 6: Quantification of discovery capability of the lemmatizer

	Accuracy
Lindén 2008	68.2%
MBSMA-s	81.2%

6. Discussion and future work

To the best of our knowledge, the research results presented above describe the first attempt at building a comprehensive data-driven morphological analyzer for a Bantu language. It improves on previous rule-based approaches in terms of development time and accuracy, as well as in its ability to handle word forms for previously unseen lemmas.

We have demonstrated how this can be achieved with relatively little manual effort, and experimental results show that the method compares favourably to a meticulously designed rule-based technique, even when it is trained on the basis of its output. Defining the problem of data-driven morphological analysis on the syllable level, rather than on the character level, we furthermore showed how techniques typically designed with Indo-European language processing in mind, can be adjusted to work for Bantu languages as well.

The system shows promising results as a lexicographic tool: the lemmatizer yields encouraging results when considered as a corpus annotation tool and can therefore be considered as a useful addition to a CQP. Furthermore, the lemmatizer enables the discovery of previously unrecorded lemmas and can also function as a component in an interface for dictionary consultation.

The performance of the system can undoubtedly still be improved. The current system has been trained on an automatically annotated corpus of Swa-

hili. While the experimental results show that the machine-learning algorithm is to some extent able to filter out the noise in the data, we believe that cleaner training data can significantly improve the accuracy of the induced morphological analyzer.

During the development of the morphological analyzer, we have made some important and necessary abstractions. In the context of building a lemmatizer, it is not problematic to limit the system to recognize entire prefix and suffix groups. A true morphological analyzer should however also be able to segment and label the individual affixes. To this end, the construction of a large Swahili morphological database would be welcome, similar in scope to CELEX. We are confident that the morphological analyzer described in this article can significantly aid the construction of this data by providing a fast and accurate automatic pre-processor to the manual annotation.

Demonstration system and acknowledgements

A demonstration system for the MBSMA-s system can be found on the AfLaT website (<http://aflat.org/?q=node/241>).

Guy De Pauw is funded as a Postdoctoral Fellow of the Research Foundation – Flanders (FWO). Gilles-Maurice de Schryver would like to thank Ghent University for its continued support of his field trips to South Africa. Both authors would also like to thank Naomi Maajabu (<http://aflat.org>) for her annotation efforts.

Endnotes

1. For an online demo of the Swahili tagger, consult <http://aflat.org/?q=node/10>
2. For an online demo of the Northern Sotho tagger, consult <http://aflat.org/?q=node/177>

References

- Anderson, W.A., P.M. Kotzé and A.E. Kotzé.** 2007. *Application and Testing of Performance Enhancing Morphological Analysis Techniques*. Paper presented at the LSSA/SAALA/SAALT Joint Annual Conference, held at the North-West University, Potchefstroom, Republic of South Africa, 4–6 July 2007.
- Baayen, R.H., R. Piepenbrock and L. Gulikers.** 1995. *The Celex Lexical Database (Release2) [CD-ROM]*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Bosch, S.E., J. Jones, L. Pretorius and W.A. Anderson.** 2006. Resource Development for South African Bantu Languages: Computational Morphological Analysers and Machine-Readable Lexicons. Roux, J. (Ed.). 2006. *Proceedings of the Workshop on Networking the Development of Language Resources for African Languages, LREC 2006*: 38–43. Genoa: ELRA.

- Christ, O.** 1994. A Modular and Flexible Architecture for an Integrated Corpus Query System. *Proceedings of COMPLEX'94: 3rd Conference on Computational Lexicography and Text Research*: 23-32. Budapest: Hungarian Academy of Sciences, Linguistics Institute.
- Creutz, M., K. Lagus, K. Lindén and S. Virpioja.** 2005. Morfessor and Hutmegs: Unsupervised Morpheme Segmentation for Highly-Inflecting and Compounding Languages. Langemets, M. and P. Penjam (Eds.). 2005. *Proceedings of the Second Baltic Conference on Human Language Technologies*: 107-112. Tallinn: Tallinn University of Technology.
- De Pauw, G., G.-M. de Schryver and P.W. Wagacha.** 2006. Data-driven Part-of-speech Tagging of Kiswahili. Sojka, P. et al. (Eds.). 2006. *Proceedings of Text, Speech and Dialogue, 9th International Conference*: 197-204. Berlin: Springer.
- De Pauw, G., T. Laureys, W. Daelemans and H. van Hamme.** 2004. A Comparison of Two Different Approaches to Morphological Analysis of Dutch. *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology*: 62-69. Barcelona: ACL.
- De Pauw, G. and P.W. Wagacha.** 2007. Bootstrapping Morphological Analysis of Gikūyū Using Unsupervised Maximum Entropy Learning. *Proceedings of the Eighth INTERSPEECH Conference, Antwerp, Belgium*.
- De Pauw, G., P.W. Wagacha and D.A. Abade.** 2007. Unsupervised Induction of Dholuo Word Classes using Maximum Entropy Learning. Getao, K. and E. Omwenga (Eds.). 2007. *Proceedings of the 1st International Conference in Computer Science and ICT*: 139-143. Nairobi: University of Nairobi.
- De Schryver, G.-M. and G. De Pauw.** 2007. Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*. *Lexikos* 17: 226-246.
- Elwell, R.** 2006. Finite State Methods for Bantu Verb Morphology. *Proceedings of the Texas Linguistics Society X, Austin*.
- Elwell, R.** 2008. Using Syllables as Features in Morpheme Tagging in Swahili. *Proceedings of the Fifth Midwest Computational Linguistics Colloquium, East Lansing*.
- Gelbukh, A. (Ed.).** 2008. *Computational Linguistics and Intelligent Text Processing: 9th International Conference, CICLING 2008 Proceedings*. Berlin: Springer.
- Goldsmith, J.** 2001. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics* 27: 153-198.
- Google.** 2008. Google Search Engine [online]. <http://google.com/>.
- Hillewaert, S., P. Joffe and G.-M. de Schryver.** 2008. *Kamusi ya Kiswahili — Kiingereza Katika Mta-ndao/ Online Swahili-English Dictionary* [online]. <http://africanlanguages.com/swahili/>.
- Hurskainen, A.** 1992. A Two-Level Computer Formalism for the Analysis of Bantu Morphology: An Application to Swahili. *Nordic Journal of African Studies* 1: 87-119.
- Hurskainen, A.** 1996. Disambiguation of Morphological Analysis in Bantu Languages. *Proceedings of the 16th Conference on Computational Linguistics*: 568-573. Copenhagen: ACL.
- Hurskainen, A.** 2004. Swahili Language Manager: A Storehouse for Developing Multiple Computational Applications. *Nordic Journal of African Studies* 13: 363-397.
- Hurskainen, A.** 2004a. HCS 2004 — Helsinki Corpus of Swahili. Compilers: Institute for Asian and African Studies (University of Helsinki) and CSC.
- Hurskainen, A. and R. Halme.** 2001. Mapping Between Disjoining and Conjoining Writing Systems in Bantu Languages: Implementation on Kwanyama. *Nordic Journal of African Studies* 10: 399-414.

- Kamusi Project*. 2008. The Internet Living Swahili Dictionary [online]. <http://kamusiproject.org/>.
- Kilgarriff, A. et al.** 2008. *Sketch Engine* [online]. <http://sketchengine.co.uk/>.
- Koskenniemi, K.** 1983. *Two-level Morphology: A General Computational Model for Word-form Recognition and Production*. Helsinki: University of Helsinki, Department of General Linguistics.
- Kotzé, P.M. and W.A. Anderson.** 2005. A Computational Morphological Analyser for Northern Sotho Deverbative Nouns: Applying Xerox Finite-state Software to Traditional Grammar. *South African Journal of African Languages* 25: 59-70.
- Lindén, K.** 2008. A Probabilistic Model for Guessing Base Forms of New Words by Analogy. Gelbukh, A. (Ed.). 2008: 106-116.
- Maphosa, M.** 2002. *Word Division and Orthography as Some of the Factors Posing Challenges in the Development of the Ndebele Grammatical Parser*. Paper presented at the Seventh International Conference of the African Association for Lexicography, organized by the Dictionary Unit of South African English, Rhodes University, Grahamstown, Republic of South Africa, 8–10 July 2002. For an abstract, see the AFRILEX 2002 conference brochure, 23-24.
- Masaaki, N.** 1999. A Japanese Morphological Analysis Method Using a Statistical Language Model and an *N*-best Search Algorithm. *Transactions of Information Processing Society of Japan* 40: 3420-3431.
- Muhirwe, J.** 2007. Computational Analysis of Kinyarwanda Morphology: The Morphological Alternations. Kizza, J.M. et al. (Eds.). 2007. *Special Topics in Computing and ICT Research: Strengthening the Role of ICT in Development*: 78-87. Kampala: Fountain Publishers.
- Ngugi, K., W. Okelo-Odongo and P.W. Wagacha.** 2005. Swahili Text-to-Speech System. *African Journal of Science and Technology* 6(1): 80-89.
- Oflazer, K.** 2008. Statistical Machine Translation into a Morphologically Complex Language. Gelbukh, A. (Ed.). 2008: 376-388.
- Pretorius, L. and S.E. Bosch.** 2007. Containing Overgeneration in Zulu Computational Morphology. Vetulani, Z. (Ed.). 2007. *Human Language Technologies as a Challenge for Computer Science and Linguistics, Proceedings of 3rd Language and Technology Conference*: 54-58. Poznan: Wydawnictwo Poznańskie Sp. z o.o.
- Ridings, D. and W. Mavhu.** 2002. *Problems and Challenges Encountered when Developing a Morphological Parser for the Shona Language*. Paper presented at the Seventh International Conference of the African Association for Lexicography, organized by the Dictionary Unit of South African English, Rhodes University, Grahamstown, Republic of South Africa, 8–10 July 2002. For an abstract, see the AFRILEX 2002 conference brochure, 24-26.
- Taljar, E. and S.E. Bosch.** 2005. A Comparison of Approaches Towards Word Class Tagging: Disjunctively vs Conjunctively Written Bantu Languages. *Proceedings of the Conference on Lesser Used Languages and Computer Linguistics, Bolzano*.
- Van den Bosch, A. and W. Daelemans.** 1999. Memory-based Morphological Analysis. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*: 285-292. Maryland: ACL.
- Van den Bosch, A., E. Marsi and A. Soudi.** 2007. Memory-based Morphological Analysis and Part-of-speech Tagging of Arabic. Soudi, A. et al. (Eds.). 2007. *Arabic Computational Morphology: Knowledge-based and Empirical Methods*: 203-219. Berlin: Springer.

Two Recent Major Afrikaans–English/English–Afrikaans Dictionaries from Pharos

Fred Pfeiffer, *Pharos Dictionaries, Cape Town* (pfeiffer@pharos.co.za)

Abstract: When Pharos Dictionaries was established in 1996, its first order of business was to develop a comprehensive Afrikaans–English/English–Afrikaans dictionary that could succeed the standard-bearing but ageing TW (*Tweetalige Woordeboek/Bilingual Dictionary* by Bosman, Van der Merwe and Hiemstra). The article explains in detail and with examples how this new dictionary, *Pharos One* for short, was developed from the TW. Information about syllabification, grammar and usage were greatly enhanced, the lemma structure was made more accessible and thousands of new words and expressions were added. The most difficult aspect of the development process was the incorporation of older material from the TW. For the sake of comprehensiveness much of this material was retained, but because of the sensitive nature of many words and expressions from the old South Africa, usage instructions had to be added with great care. For the editing of the English–Afrikaans half the Pharos editors could draw on a plethora of quality British and American and bilingual European dictionaries, but for the Afrikaans–English half the editors had to rely on their own language knowledge. The article explains how with reference to electronic archives and the help of outside consultants the Pharos team could verify their judgment. In response to lagging sales, Pharos embarks on the development of a scaled-down version of *Pharos One* that would retain its best characteristics, but would have a more manageable size and affordable price. The article explains how a one-third reduction of *Pharos One* was achieved to produce this new dictionary, called the *Concise* for short.

Keywords: ALPHABETISATION, COMPOUND, COMPREHENSIVENESS, CROSS-REFERENCING, DATABASE, DATED MATERIAL, DERIVATIVE, DICTIONARY USER, ENTRY, EXAMPLE PHRASE, FREQUENCY COUNTS, HOMOGRAPH, IDIOMATIC EXPRESSION, LEMMA, OFFENSIVE WORD/EXPRESSION, SYLLABIFICATION, TECHNICAL TERMINOLOGY, TRANSLATION, USAGE INSTRUCTION

Opsomming: Twee onlangse groot Afrikaans–Engels/Engels–Afrikaanse woordeboeke van Pharos. Toe Pharos Woordeboeke in 1996 tot stand gebring is, was sy eerste taak om 'n omvattende Afrikaans–Engels/Engels–Afrikaanse woordeboek saam te stel wat die toonaangewende dog verouderende TW (*Tweetalige Woordeboek/Bilingual Dictionary* deur Bosman, Van der Merwe en Hiemstra) kon vervang. Hierdie artikel verduidelik in besonderhede en met voorbeelde hoe dié nuwe woordeboek, kortweg *Pharos Een*, uit die TW ontwikkel is. Inligting oor sillabeverdeling, grammatika en gebruik is sterk uitgebrei, die lemmastruktuur is toegankliker gemaak en duisende nuwe woorde en uitdrukkings is bygevoeg. Die moeilikste deel van die ontwikkelingsproses was die inkorporering van ouer materiaal uit die TW. Ter wille van omvattendheid is baie van dié materiaal behou, maar weens die sensitiewe aard van baie woorde en uitdrukkings uit die ou Suid-Afrika, moes gebruikstoeliggings met groot omsigtigheid toegevoeg word. Vir

die redigering van die Engels–Afrikaanse gedeelte kon die Pharosredakteurs uit 'n groot verskeidenheid Britse en Amerikaanse en tweetalige Europese woordeboeke put, maar vir die Afrikaans–Engelse gedeelte was die redakteurs op hul eie taalkennis aangewese. Die artikel verduidelik hoe die Pharosspan hul eie oordeel met verwysing na elektroniese argiewe en met die hulp van buiteraadgewers kon kontroleer. In reaksie op die trae verkope onderneem Pharos om 'n beknopter weergawe van *Pharos Een* te ontwikkel wat sy beste eienskappe sou behou, maar van 'n meer hanteerbare grootte en bekostigbare prys sou wees. Die artikel verduidelik hoe 'n reduksie van een derde van *Pharos Een* bereik is om dié nuwe woordeboek, kortweg die *Kernwoordeboek*, te lewer.

Sleutelwoorde: AFLEIDING, ALFABETISERING, DATABASIS, FREKWENSIETELLINGS, GEBRUIKSLEIDING, HOMOGRAAF, IDIOMATIESE UITDRUKKING, INSKRYWING, KRUIS-VERWYSING, KWETSENDE WOORD/UITDRUKKING, LEMMA, OMVATTENDHEID, SAMESTELLING, SILLABEVERDELING, TEGNIESE TERMINOLOGIE, VEROUDERDE MATERIAAL, VERTALING, VOORBEELDFRASE, WOORDEBOEKGEBRUIKER

1. A new South African dictionary publisher

Pharos Dictionaries was founded in 1996 to amalgamate the various dictionary titles from Naspers. Dictionaries hitherto published by Tafelberg and J.L. van Schaik, for example, would from now on bear the Pharos imprint. The company would also provide a home for those editors with a lexicographical bent to ply their trade. With manager Hans Büttner, and with editors Madaleine du Plessis and Jana Luther to tackle the text work, Pharos Dictionaries set about modernising the titles it had inherited, but also preparing for the publication of brand new titles.

A thorough revision of Bosman, Van der Merwe and Hiemstra's *Tweetalige Woordeboek/Bilingual Dictionary* (the TW) was the fledgling company's most immediate and pressing task. Generally considered the standard, authoritative reference work for translating English to Afrikaans and vice versa, the TW was increasingly falling behind the requirements of language practitioners in particular and, in general, anyone who regularly had to write clearly, correctly and idiomatically in both Afrikaans and English. The last revision of the TW, the eighth (TW8), by Joubert and Spies, was completed in 1984 and with the establishment of Pharos Dictionaries that was already 12 years ago. In dictionary terms, that is a fairly lengthy period. Large overseas dictionary publishers like Oxford and Collins seem to manage at least two revisions of their frontline titles every decade.

A further complicating factor was that the years that had lapsed since the publication of TW8 were no ordinary years. During this period the Berlin Wall had fallen and here at home South Africans had their first fully democratic elections. Both these events brought about a complete reordering of the social, political and geographical structure, globally and locally. On the technological front these were equally revolutionary times. Personal computers became commonplace, and the Internet and cell phones opened up a whole new world. It was also the era in which the shadow of Aids began to creep across the globe

and phenomena like the greenhouse effect began to make their presence felt. Tumultuous developments such as these find their expression in language almost simultaneously, and within a reasonable time dictionary users expect to find in their favourite dictionaries the new vocabulary generated in this way.

Work on the revision of TW8 started in earnest soon after the establishment of Pharos Dictionaries. Madaleine du Plessis was chief editor, and two more editors were appointed for the task, namely Wanda Smith-Müller and Fred Pheiffer. Although the revision of TW8 would consume most of Pharos's time and energy over the next couple of years, this was by no means its only project. Its first two publications were both aimed at meeting the demand for an up-to-date Afrikaans–English/English–Afrikaans (AEEA) dictionary. The 14th edition of the *Groot Woordeboek/Major Dictionary* edited by Louis Eksteen, a title inherited from J.L. van Schaik, provided some improvement on TW8, but was not sufficiently modernised to replace it and ultimately fell short of expectations. The *Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary* by Phillip Joubert, a title inherited from Tafelberg, proved to be very useful, but as a specialised dictionary it was intended at best to augment the TW.

Pharos's first 'home-grown' publication was *New Words/Nuwe Woorde* by Madaleine du Plessis in 1999. It contained Afrikaans translations for well over 10 000 modern English words and expressions, material hitherto not covered by the TW, but also new, additional meanings and uses of terms already contained in the TW. This publication was the first product of the revision work being done on TW8 and as such provided a sample of what the successor to the TW would offer. It contained words used in a wide range of disciplines and words reflecting technological advances, but also many new phrases and idioms and a good number of informal and slang words and expressions. However, as Joubert's *Phrase Dictionary*, it was designed to be used as a supplement to the TW, not as a standalone dictionary.

Although all three these dictionaries were ultimately superseded by or assimilated into the new AEEA dictionary, they all proved their worth as vital stand-ins during the gap between the TW and its successor, a period that eventually stretched to two decades. Moreover, they became essential cogs in Pharos's first electronic product, the *Pharos 5 in 1 Dictionaries/Woordeboeke* CD-ROM, a vehicle that prolonged their usefulness. It was first released in 2000 and is still proving popular in 2008.

2. A successor to the TW

The development of a new top-of-the-range AEEA dictionary to succeed the TW encompassed the following additions and alterations to the text of TW8: inserting syllable separators and underlining the stressed syllable in all headwords and their derivatives; restructuring many of the entries; shuffling the order of the translations to place current meanings first, and older or specialised meanings further back; adding new material; and treating dated material.

2.1 Syllable treatment

Below is a comparative entry from TW8 and *Pharos One*, showing the addition of syllable separator dots and stress underlining in both the main entry and the derivatives in the latter:

TW8: **Nige´rië** Nigeria. **Nige´riër** =s, n. Nigerian. **Nige´ries** =e, a. Nigerian.

Pharos One: **Ni·ge·ri·ë** (*geog.*) Nigeria. **Ni·ge·ri·ër** =riërs, n. Nigerian. **Ni·ge·ries**
=riese, *adj.* Nigerian.

Note also the expansion of the grammatical and usage information in *Pharos One*.

Syllabification was generally easier to effect on the Afrikaans–English (AE) half than the English–Afrikaans (EA) half of the dictionary, since Afrikaans syllabification is phonetic, and to apply it is relatively straightforward for anyone with a sound knowledge of spoken Afrikaans. English syllabification on the other hand is morphological and requires knowledge of the underlying rules or, alternatively, reference to sources with English syllabification. Surprisingly, British English dictionaries generally do not contain syllabification. The 1985 edition of the *Collins Dictionary of the English Language* seems to have been the last to offer comprehensive information on it. Bilingual European dictionaries with an English component like those of *Van Dale* and *Langenscheidt* show syllabification and American English dictionaries like the *American Heritage* and the online versions of *Merriam-Webster* and *Encarta* all have syllabified lemmas.

Intriguingly, the syllabifications offered in these sources do not always match. The table below shows the differing treatment of a few randomly picked examples from the above-mentioned edition of *Collins* and the 1993 edition of *The American Heritage College Dictionary*. The difference in interpretation seems to be influenced by variation in the placement of the stressed syllable (for example, both *homophyly* and *homophyly* are acceptable) but the two sources do not offer different syllabifications for the variant pronunciations.

	<i>Collins</i>	<i>American Heritage</i>
demonstrator	de·mon·stra·tor	dem·on·stra·tor
dirigible	di·rig·i·ble	dir·i·gi·ble
disinfectant	dis·in·fect·ant	dis·in·fec·tant
homophyly	ho·moph·y·ly	ho·mo·phy·ly
serotonin	ser·o·to·nin	se·ro·to·nin
tachisme	ta·chisme	tach·isme
tergiversate	ter·gi·versate	ter·giv·er·sate
uropygium	u·ro·pyg·i·um	u·ro·py·gi·um
volcanology	vol·can·ol·o·gy	vol·ca·nol·o·gy

Admittedly, these are all fairly low-frequency technical words, except maybe for *demonstrator* and *disinfectant*, and the level of agreement in the above-mentioned sources is probably close to 99%, but in the quest for preciseness in one's own dictionary discrepancies in the sources one relies on can be frustrating.

2.2 Restructuring

Many of the entries in TW8 tend to be very long and overly complex, because of an attempt to place as many derivatives and compounds under a single lemma as possible. This type of lemma organisation may be highly systematic, but it is not necessarily user-friendly. In an effort to make the information in the entries as accessible as possible, such long entries had to be disentangled. In particular, wherever derivatives of the main entry were placed among compounds of the same entry, they were removed and elevated to separate main entries. Hence derivatives and compounds were not allowed to coexist in the same main entry (except of course where the derivatives followed after the compounds, alphabetically speaking). Below are comparative entries from TW8 and *Pharos One*; in the latter derivatives of **ego** have been made separate entries. The main advantage is that the reference of the compounds' tilde to the main entry is apparent, without interference from derived forms.

TW8: **e'go** *egos* ego, ek, self; ekheid, eie ek. **egocen'tric** egosentries, selfgerig, in sigself opgaande. **egocentri'city** egosentrisiteit. **e'goism** egoïsme, selfsug, selfliefde. **e'goist** egoïs, selfsugtige. **egoist'ic** egoïsties, selfsugtig. **egoman'ia** uiterste selfsug; sieklike egoïsme. **egoman'iac** sieklike egoïs. **e'gotism** selfingenomenheid, eiedunk, eiewaan, eieliefde, egotisme. **egotist'ic(al)** selfingenome, in sigself opgaande, selfbehep, ekkerig, egotisties. ~ **trip** selfverheerliking.

Pharos One: **e-go** =*gos* ego, ek, self; ekheid, eie ek; *boost / feed one's* ~ jou ego streef, jou eiedunk verhoog. ~ **ideal** (*psig.*) ego-ideaal. ~ **identity** (*psig.*) ego-identiteit. ~ **psychology** egopsigologie, =sielkunde. ~ **trip** selfverheerliking; *be on an ~ ~, (infnl.)* met selfverheerliking besig wees.
e-go-cen-tric egosentries, selfgerig, in jousef opgaande. **e-go-cen-tric-i-ty** egosentrisiteit.
e-go-ism egoïsme, selfsug, selfliefde. **e-go-ist** egoïs, egosentriese / egoïstiese / (eie-)ekkerige / selfbehepte mens, selfsugtige (mens); selfingenome / selfvoldane / verwaande mens; (*etiek*) egoïs. **e-go-ïs-tic**, **e-go-ïs-ti-cal** egoïsties, selfsugtig.
E-go-li, **e-Go-li** (*Z.: Johannesburg*) Goudstad.
e-go-ma-ni-a uiterste selfsug; sieklike egoïsme **e-go-ma-ni-ac** sieklike egoïs.
e-go-tism selfingenomenheid, selfvoldaanheid, eiedunk, eiewaan, eieliefde, egotisme. **e-go-tist** selfingenome / selfvoldane / verwaande mens; egoïs, egosentriese / egoïstiese / (eie-)ekkerige / selfbehepte mens, selfsugtige (mens). **e-go-tis-tic**, **e-go-tis-ti-cal** selfingenome, selfvoldaan, verwaand, egotisties; egosentries, egoïsties, (eie-)ekkerig, selfbehep, selfsugtig.

Note also the expanded use of usage labels and example phrases in the *Pharos One* excerpt.

Performing such disentanglement procedures manually (at this stage of the editorial process cutting and pasting in XYWrite, an obsolete yet very stable word-processing program) posed at least two risks. Firstly, derivatives could be removed from an entry and posted as new entries immediately after their mother entry without checking whether this procedure disturbed the alphabetical order of the entries that previously followed immediately after the mother entry. A second risk is that in the cutting and pasting of derivatives one could get distracted and forget to paste. In TW8 the lemma **suit** (p. 1235) has the derivatives **suitability**, **suitable**, **suitably**, **suited**, **suiting** and **suitor** dispersed among the compounds **suitcase**, **suitcase cover** and **suit length**, but in *Pharos One* only **suited** and **suiting** survived (p. 1407), which is most likely an oversight of the kind described above. The first risk can be obviated by regularly running a script over the lemma list to check for alphabetical errors, but the second risk is more serious. A dedicated lexicographical program like TshwaneLex, which *Pharos Dictionaries* started using in 2006, probably offers the best solution. Not only does it alphabetise automatically, but also alerts the user whenever deletion of any part of a lemma is attempted.

2.3 Reordering translations

Any revision or renewal of the semantic content of the lemmas in TW8 relied in the first instance on the lexicographer's own sense of language, on a hunch, admittedly, that the translations given in TW8 did not quite match the current usage of Afrikaans or English in South Africa. Although constantly aware that such an approach was perhaps not best practice, the constraints of time left one with very little other option. Fortunately, with the revision of the EA half of TW8, which was tackled first, one had the luxury of being able to refer to several first-class, up-to-date English dictionaries. In cases where one felt TW8 had become dated or deficient, one could compare its translations with the meanings offered by *Collins* or *Oxford*, for example. Such a comparison showed that TW8's content was basically sound, but confirmed that it had dated in several respects. Below are two examples comparing entries from TW8 and *Pharos One*, which show how the focus of the meaning of the words concerned has shifted.

In the first example the translation of **reduction** by 'terugbrenging', which is unlikely to find much resonance nowadays, has been pushed right back in *Pharos One*, allowing the expected translation of 'vermindering' etc. to stand first. Note the use of the label (*arg.*) to indicate that certain translations have become archaic. Further note how *Pharos One* places usage labels like (*wisk.*) and (*med.*) before a translation, where they can refer to both the lemma and its translation.

TW8: **reduc'tion** terugbrenging; vermindering, verkleining, inkrimping, inkorting; besnoeiing; afname; verlaging; herleiding (*mat.*); korting, afslag, rabat; setting (*chir.*); degradering (*in rang*); onderwerping; inneming, verowering; reduksie (*chem.*); verswakking (*tot.*); ~ *to normal* normalisering. ~ **division** reduksiedeling. ~ **roller** reduksieroller. ~ **worker** reduksiewerker. ~-**works** reduksiewerkplaas, smeltoond.

Pharos One: **re-duc-tion** vermindering, verkleining, inkrimping, inkorting; besnoeiing; afname; verlaging; korting, afslag, rabat; terugbrenging; (*wisk.*) herleiding; (*med.*) setting; (*chem.*) reduksie; (*tot.*) verswakking; degradering (*in rang*); (*arg.*) onderwerping; (*arg.*) inneming, verowering; *a ~ to normal* normalisering; *sweeping ~s* algemene / kolossale / reusagtige prysverlagings. ~ **division** reduksiedeling. ~ **roller** reduksieroller. ~ **worker** reduksiewerker. ~ **works** reduksiewerkplaas, smeltoond.

In the second example, the usual translation of **uitskel** one would nowadays expect, namely 'scold', is buried deep in the TW8 entry. In the *Pharos One* entry, 'scold' has been elevated to the foremost position and 'abuse' has been modified by 'verbally'. The example phrase *iemand uitskel* has been extended considerably with idiomatic translations, and the name-calling aspect has been relegated to the back of queue. Note how the scope of the second example phrase has been broadened in *Pharos One* by the use of the ellipsis and the offering of an example (*liar etc.*) as a possible complement.

TW8: **uit'skel** *uitge*= abuse, call names, swear at, slate, berate, inveigh against, chide, slang, 'bawl out', scold, revile; *iem.* ~ call s.o. names; *iem.* ~ *vir 'n skelm* brand s.o. as a rogue. **uit'skellery** abuse, abusiveness, invective, slanging-match. ~**party** slanging-match. ~**woord** term of abuse.

Pharos One: **uit-skel** *uitge*= scold (severely), abuse (verbally), call names, swear at, slang, bawl out, chide (severely), slate, berate, inveigh against, revile; *iem.* ~ give s.o. the rough / sharp edge / side of one's tongue, tongue-lash (*or* lash into) s.o., (*infml.*) give s.o. what for (*or* a tongue-lashing), (*infml.*) bawl s.o. out; call s.o. names; *iem.* *vir 'n ...* ~ call s.o. (*or* brand s.o. as) *... (liar etc.)*. ~**party** slanging match. ~**woord** term of abuse. **uit-skel-le-ry** abuse, abusiveness, invective, slanging match.

Compared to the EA half, the editing of the AE half had to proceed without fresh Afrikaans lexicographical resources. Pharos's own *Verklarende Afrikaanse Woordeboek* (the VAW) predated 1994 — that benchmark date for all modern South African reference works. A new edition of the *Verklarende Handwoordeboek van die Afrikaanse Taal* (the HAT) was published in 2000, but it was not vastly different from earlier editions. Of the standard reference work of the Afrikaans language, the *Woordeboek van die Afrikaanse Taal* (the WAT), less than two-thirds of the letters had been done (the letters N to O were completed during the editing period of *Pharos One*). Moreover, many of the WAT's earlier volumes had become severely outdated and as a result contained the very flaws that one sought solutions for in order to produce *Pharos One*.

The implication of this paucity of suitable references was that one had to rely — even more so than in the EA half — on one's own knowledge of language, in particular of Afrikaans in its present-day form. *Pharos* did have at its disposal two types of corrective or control though. The first of these are the Media24 newspaper and magazine archives and the archive *Pharos* has built up from manuscripts from its sister companies like Tafelberg and Human & Rousseau and from other material. These electronic databases provide a means of checking the frequency, usage and most current meaning of words and expressions. And although gleaning information in this way is probably the ideal approach to lexicography, working with raw data is a laborious process. Because time was of the essence, this research method could only be resorted to when one's own ability to judge with certainty the validity or correctness of a TW8 entry failed.

The second corrective was to employ the services of a number of seasoned Afrikaans linguists (Fritz Ponelis, Ernst Kotzé, Tom McLachlan and the late Johan Combrink) and editors (Louise Voigt and George Boshoff) to read and comment on the *Pharos One* manuscript in the various stages of its preparation. As outsiders not intimately involved in the project, they could provide input that was objective and valued for its authoritativeness.

2.4 New material

The bulk of the new material in *Pharos One* came straight from *New Words/Nuwe Woorde*. It was added to the EA half of *Pharos One* (*New Words* being one-directional EA only) either as new lemmas or integrated into existing entries as new meanings, expressions, compounds or derivatives. In the example below, the top excerpt is the original section from TW8 and the middle one from *New Words* while the bottom excerpt shows the amalgamation in *Pharos One*. In the excerpt from *New Words*, **Fosbury flop** provides *Pharos One* with an altogether new entry, **fossil fuel** is a new compound to an existing noun entry, while **fossilised** is the adjectival use of the past participle of an existing verb.

TW8: **forwear'ied, forworn'** uitgeput.
fos'korite foskoriet.
fosse grag, sloot.
fos'sick snuffel, soek; rondpeuter; prospekter.
fos'sil *n.* fossiel. **fos'sil** *a.* fossiel, versteen(d) (*lett.*); verstar(d), versteen(d) (*fig.*). ~ **fish** vissteen. **fossili'ferous** fossielhoudend. ~ **impression** fossielafdruk. **fossilisa'tion, =za'tion** fossilering; verstarling, verstering (*fig.*). **fos'silise, =ize** fossileer; (laat) verstar of versteen (*fig.*).
fos'silist fossiel(e)kenner. ~ **plant** fossiele of gefossileerde plant.
 ~ **topography** fossiele topografie.

New Words: **Fos-bur-y flop** *n.*, (*hoogspring*) fosburystyl.
fos-sil *n.* & *a.*: ~ **fuel** fossielbrandstof.
fos-sil-ised, =ized *a.*, (*paleont.* & *fig.*) versteen(d) (*dier, taalvorm*); uitge-

dien(d) (*gebruike*); onbuigsaam, onversetlik, verstok (*iem.*); ~ *burocrat* uitgedroogde burokraat; ~ *remains* fossiele.

Pharos One: **for-wea-ried, for-worn** (*arg.*) uitgeput.
Fos-bur-y flop (*atl.*) fosburystyl.
foss(e) (*argeol.*) grag, sloot.
fos-sick (*Austr., infml.*) snuffel, soek; rondpeuter; prospekter.
fos-sil *n.*, (*geol. of neerh.*) fossiel. **fos-sil bep.**, (*geol.*) fossiel-, versteende; (*fig., neerh.*) verkrampte, verstokte, verstarde. ~ **fish** vissteen. ~ **fuel** fossielbrandstof. ~ **impression** fossielafdruk. ~ **plant** fossielplant, versteende / gefossileerde plant. ~ **topography** fossiele topografie.
fos-sil-if-er-ous (*geol.*) fossielhoudend.
fos-sil-ise, =ize fossileer, versteen; (*fig.*) (laat) verstar / versteen.
fos-sil-i-sa-tion, =za-tion fossilering, verstening; (*fig.*) verstarring, verstening. **fos-sil-ised, =ized** (*paleont. of fig.*) versteen(d) (*dier, taal= vorm*); uitgedien(d) (*gebruike*); onbuigsaam, onversetlik, verstok (*iem.*); ~ *bureaucrat* uitgedroogde burokraat; ~ *remains* fossiele.
fos-sil-ist fossielkenner.

Note once again how the derivatives in the **fossil** entry in TW8 have been removed and elevated into independent entries in the *Pharos One* excerpt to allow the compounds of **fossil** to be closer to and follow uninterrupted after the lemma. Alphabetisation in the new **fossilise** entry has been suspended so that **fossilisation** can be listed as a derivative of the lemma.

Note too once again the expanded use of usage labels in the *Pharos One* excerpt. The approach throughout to apparently dated or rare material in TW8 was not to summarily dismiss such entries, but to try to accommodate as many of them as possible through the clever application of usage labels. Hence (*arg.*), (*argeol.*) and (*Austr., infml.*) to indicate archaic, archaeology and Australian English or informal. However, an example like **foskorite**, which has disappeared from *Pharos One*, would have been unknown to the editor working on that section of TW8, would furthermore not have been found in any of the British, American, Dutch or German dictionaries in the Pharos office, including a venerable edition of *Webster's*, and finally, would in all likelihood have returned a low Google search score (less than one hundred).

Where new material in the EA half of *Pharos One* gave rise to new Afrikaans translations too, a 'mirror image' of the entry was created allowing the translations to be used as new lemmas in the AE half. In the excerpts below, the new compounds of **gift** in *New Words* gave rise to the following new compounds of **geskenk** in *Pharos One*: **geskenkpak(kie)**, **geskenkdoos**, **geskenkbelasting** and **geskenkpapier**.

TW8: **geskenk** =e gift, present; offering, donation; *ten ~e bied (gee)* make a present of, give as a present, present with. **-bewys** gift voucher.

New Words:: **gift** *n.*: ~ **box**, ~ **pack** geskenkpak(kie), =doos. ~ **tax** geskenkbelasting, belasting op geskenke. ~ **wrap**, ~ **wrapping** *n.* geskenkpapier. **~wrap** *v.* in / met geskenkpapier toedraai.

Pharos One: **ge-skenk** =skenke, n. gift, present, (*infml.*) pressie, prezzie; offering, donation; *iem. iets ten ~e bied / gee*, (*obs., fml.*) make s.o. a present of s.t., give s.o. s.t. as a present, present s.o. with s.t.; 'n ~ vir ... van ... a present for ... from ... **ge-skenk** =skenkte, *adj.* given, granted; donated. **~belasting** gift tax. **~bewys** gift voucher. **~eksemplaar** =plare gift copy, giveaway copy (of a book etc.). **~pak(kie)**, **~doos** gift box / pack. **~papier** gift wrap(ping); *iets in / met ~ toedraai* giftwrap s.t..

Note how the example phrase and its translations in the TW8 excerpt have been expanded in *Pharos One* to include both subject and object. This is largely the legacy of the work done by Joubert in his *Bilingual Phrase Dictionary*, from which *Pharos One* borrowed extensively.

Other sources of new English material were *The New Oxford Dictionary of English*, particularly the *Concise Oxford Dictionary*, and the *Collins English Dictionary*. A relatively late source of new words was the seventh edition of *Roberts Birds of Southern Africa*. In this edition, a number of the local English bird names were changed to make them correspond with international English bird names (*lapwing* instead of *plover*, *turaco* instead of *loerie*, *thick-knee* instead of *dikkop*, etc.). As many of these changes as possible were implemented in *Pharos One*, but because the letters were typeset as they were completed, this was not always practical. Nevertheless, *Pharos One* was probably one of the first reference works to contain the new nomenclature from *Roberts*.

2.5 Dated material

As described in the previous section, the editing policy applied during the preparation of *Pharos One* was to preserve as much of the dated material as possible, rather than get rid of it wholesale, which would have been the knee-jerk approach. However, preserving dated material for the sake of comprehensiveness implied a responsibility to the dictionary user to identify such material with care and to label it accordingly. Dated material was classified as obsolete (old-fashioned and no longer current), archaic (language relics used occasionally for effect), rare (not current, but not necessarily dated) or historic (the object referred to is dated, but not necessarily the word). Obviously these categories overlapped and labelling dated material one way or the other was not a straightforward task. However, as mentioned above, having established English dictionaries as references during the preparation of the EA half was a great boon. Especially the *New Oxford* and the *Concise Oxford* are profuse in their labelling. For the identification and labelling of dated material in the AE half, the same practice was followed as described in section 2.3.

Given the history of racial oppression in South Africa, great care had to be taken to isolate terms and expressions that still reflected racism. Many blatantly racist words had already been labelled as derogatory in TW8 (see for example pp. 235 and 964) and in *Pharos One* this process was continued with rigour. For example, for each of the myriad Afrikaans plant or animal names that is a

compound of either *hottentot* or *kaffer* a non-offensive alternative was found (see pp. 230 and 256). Words such as *Bantoe/Bantu* (TW8, pp. 46 and 710; *Pharos One*, pp. 44 and 796), *Kleurling/Coloured* (TW8, pp. 254 and 769; *Pharos One*, pp. 276 and 865) and *naturel/native* (TW8, pp. 338 and 1036; *Pharos One*, pp. 368 and 1179) that were still deemed inoffensive in the days of the TW, or at least not marked in it as problematic, were labelled as derogatory and as obsolete or historic in *Pharos One*.

A certain group of Afrikaans idiomatic expressions proved quite vexing to categorise and label. These are the expressions with corporal punishment as theme. Examples from *Pharos One*: *iem. op sy baadjie gee* give s.o. a hiding, dust/trim/warm s.o.'s jacket; *iem. op sy bas gee* tan/whip s.o.'s hide, give s.o. a (good) hiding/licking; *'n afgedankste/deftige/gedugte pak (slae)* a sound beating; a good/sound hiding; *iem. 'n groot pak gee* beat s.o. hollow; *iem. 'n helse pak (slae) gee* beat the hell out of s.o.; *iem. ('n pak/drag) slae gee* give s.o. a beating; give s.o. a hiding/tanning; *jy kan maar jou lyf vetsmeer* there's a rod in the pickle for you, prepare (yourself) for a hiding.

Although hardly derogatory, their gleeful extolling of physical violence is offensive and certainly out of step with a culture of human rights. The English translations with their emphasis on goodness and soundness are not above reproach either. Leaving them unmarked would therefore have been insensitive, but how to deal with them? Judging from their variety, these expressions still seem very much part of the Afrikaans vocabulary. As to their frequency of use, the corpora contained in the Pharos's text archives would have thrown light on that but the urgency of the work on *Pharos One* precluded a thorough investigation. However, a cursory search of the newspaper archives indicated that their use tends towards the figurative and the colloquial, and that they are often employed in sport journalism, describing rugby matches for example. This is how some of these expressions were eventually labelled, i.e. (*infnl.*) or (*sport journ.*), hardly satisfactory, but dictionary tags are quite limited when attempting to categorise such a subtle issue. In any case, it was only after the editing of the AE half was well on its way that the extent of these expressions became apparent and, as noted at the end of section 2.4, unless one had to deal with an error one tried to avoid altering pages that were already typeset.

Generally speaking though, adding new material to *Pharos One* was by far the easier operation than dealing with dated material. As the editing work on the text of TW8 progressed, a distinct impression began to take root that during the life of the TW, successive editions must have been augmented by new material, but not enough was done to evaluate existing material for its currency or appropriateness. It would therefore hardly be surprising if a study found that certain words and expressions ingrained in the TW survived from the first half of the previous century. All those Afrikaans idiomatic expressions steeped in an agrarian past are a case in point. Quaint they may have become, but do they really still speak to a modern, urbanised youth? How long can one justify retaining them in a dictionary, even for the sake of comprehensiveness?

With close to a decade of intense, dedicated and informed revision behind

it, *Pharos One* can rightly claim to be the most authoritative and up-to-date AEEA dictionary by far. Yet what the issues raised above point to is that the dictionary still carries a lot of baggage from the past which has not been unpacked and sorted. For it to become truly modern and revitalised, its contents would have to be evaluated far more rigorously against text corpora than has been possible up to now. However, this is a labour-intensive and time-consuming approach, and whether the means that such an approach requires could readily be made available in commercial publishing remains a moot point.

3. The Concise

Relatively soon after the publication of *Pharos One* at the beginning of 2005, the marketing team of NB Publishers, of which Pharos Dictionaries is an imprint, began to report that the market was daunted by the size and price of *Pharos One*. With dimensions of 21 cm x 27 cm x 5 cm and a mass of 2,5 kg, it is quite a hefty book, and for many the initial price was probably on the steep side. After having been cloistered with *Pharos One* for so long, laboriously poring over the text, it was disappointing that one's enthusiasm at its release was not immediately and generally shared. Perhaps one had hoped, somewhat naïvely and vainly, that many of the people who owned a copy of TW8 (of which more than 170 000 copies were sold over its 20-year print run) would buy a copy of *Pharos One* the moment it was released. The initial sales of *Pharos One* were certainly not poor; in fact, they matched the average monthly sales of TW8 during the time it was a Pharos imprint. But that would be comparing *Pharos One* to the TW at the end of its long and useful life — the average monthly sales of TW8 over its entire 20-year print run are about double that of *Pharos One*. The 30 000-copy first print run of TW8 sold out in just two years.

What seems to have been overlooked was that in the heyday of the TW, South Africa was officially and effectively a bilingual country. Every public office had to be able to conduct its business in both English and Afrikaans. In particular, every institution or company involved in the legal system — Parliament, the courts, academic legal faculties, legal companies — had to produce its documents in English and Afrikaans. This created a huge demand for a dictionary like the TW. Since 1994 however, and even somewhat before that, this requirement has largely fallen away. The fact that Afrikaans is no longer officially prescribed (except in provinces like the Western Cape where Afrikaans is a major language) by itself could explain the marked difference between sales of TW8 and *Pharos One*.

Although the initial sales of *Pharos One* were slower than anticipated, there was clearly a demand for a new top-of-the-range AEEA dictionary. In order to capitalise on this demand, and also to recoup the considerable investment made to deliver *Pharos One*, NB Publishers recommended to Pharos Dictionaries to produce a dictionary offering the same scope and quality as *Pharos One*, but of a more economic price and a more manageable size, a *Pharos One*

Lite essentially. Crucially, Pharos was given just two years to accomplish this. A tall order certainly, but not impossible, since the secondary purpose for which *Pharos One* was designed was to serve as a database from which smaller dictionaries could be extracted — a mother dictionary that could generate offspring — especially after the text was contained in TshwaneLex. Unfortunately, by the time Pharos Dictionaries started evaluating TshwaneLex late in 2005, the new dictionary project, which would become the *Concise*, had already progressed too far for it to be hauled into this new program. The work schedule was just too tight for a conversion process that would undoubtedly have caused time-consuming teething problems. The *Concise* was done in MS Word from the start, certainly not ideal, since processes like alphabetisation and cross-referencing had to be done manually, but it sufficed.

Because each page of *Pharos One* is made up of three columns, cutting the equivalent of one column per page would bring about a dictionary two thirds the size of *Pharos One*. A cursory investigation showed there was enough 'non-essential' material on each page to make such a cut feasible. Clearly the comprehensiveness of *Pharos One* would have to be sacrificed, the dated material from the TW, treated with such deference in *Pharos One*, being the obvious first candidate for elimination. The care with which dated material was labelled in *Pharos One* made this process fairly easy. Two further groups of words were identified that could be cut in toto, namely subject or technical terminology, and typically British, American or Australian English. However, an important proviso was that if any word in these three groups had become assimilated into everyday Afrikaans or South African English, it would be retained.

The editorial team established for the *Concise* consisted of *Pharos One* editors Jana Luther and Wanda Smith-Müller, freelance editor Celia Slater and Fred Pheiffer as final editor. Its ability to judge what should be eliminated and what retained was of critical importance for the successful and timeous completion of the *Concise* project. The editorial working conditions for the *Concise* were essentially the same as those for *Pharos One* — quality English dictionaries for the EA half, own judgement and, where time permitted, reference to electronic archives for the AE half.

Apart from reducing each page of *Pharos One* by a third, three other processes were carried out: more new words were added, errors in *Pharos One* that the editorial team had become aware of, were rectified and, where desirable or possible, the *Pharos One* lemma structure was simplified. In most respects this represented a continuation and further refinement of the work done on *Pharos One*.

3.1 Reduction, new material and corrections

The table below compares a fairly randomly chosen page from *Pharos One* with the corresponding section in the *Concise*. The lemmas run from **Nemesis** to **netting**. They are on p. 1183 in *Pharos One*, from top left to bottom right, and in the *Concise* on p. 1231, two-thirds down the left-hand column, to just over halfway

down the left-hand column on p. 1232, which is slightly less than one *Concise* page in total. This example demonstrates the successful reduction of a three-column page in *Pharos One* to a two-column page in the *Concise*.

<i>Pharos One</i> lemma list	<i>Pharos One</i> labelling	Treatment in the <i>Concise</i> (if not indicated, <i>Pharos One</i> entry left unchanged)
Nemesis	(<i>Gr. myth.</i>)	
nenta	(<i>SA, vet.</i>)	deleted
neo=	<i>comb.</i>	fragments deleted as a rule
neo-apartheid		deleted
neoclassicism		
neo-colonialism		
neocon	(<i>chiefly Am.</i>)	deleted
neoconservatism		
neodymium	(<i>chem.</i>)	
neofascism		
Neolithic	(<i>archaeol.</i>)	
neologism		derivative neology deleted
Neo-Malthusianism		deleted
—		neomycin added
neon	(<i>chem.</i>)	some compounds deleted
neonatal		
neo-Nazi		
neophilia		deleted
neophobia		deleted
neophyte		
neoplasm	(<i>med.</i>)	
neoplasty	(<i>med.</i>)	deleted
Neoplatonism	(<i>philos.</i>)	
neoprene		
neoteric	(<i>rare</i>)	deleted
Nepal	(<i>geog.</i>)	
nepenthe(s)	(<i>poet., liter.</i>)	deleted
nephalism	(<i>rare</i>)	deleted
nepheline	(<i>min.</i>)	deleted
nephelometer	(<i>chem.</i>)	deleted
nephew		derivative nephewship deleted
nephology		
nephoscope		deleted
nephralgia		deleted
nephric		deleted
nephridium	(<i>zool.</i>)	derivative nephridiopore deleted
nephrite	(<i>min.</i>)	synonym kidney stone deleted
nephritic	noun (<i>dated</i>)	only adjective retained
nephritis		
nephroid	(<i>rare</i>)	deleted
nephrolith	(<i>med., tech.</i>)	deleted
nephrology		

nephropathy		deleted
nephrosis		deleted
nephrostome	(zool.)	deleted
nephrotomy		deleted
nepotism		
Neptune	(Rom. myth., astron.)	
neptunium	(chem.)	
nerd	(infml., derog.)	
Nereid	(Gr. myth.)	
Nereus	(Gr. myth.)	
nerine	(SA, bot.)	
nerk	(Br., infml.)	deleted
neroli (oil)		
nerval		deleted
nervate	(bot.)	
nerve		several example phrases and compounds deleted
nerveless		
Nervii	(hist.)	deleted
nervine		deleted
nervous		several example phrases deleted
nervure	(entom.)	deleted
nervy	(chiefly Br., infml.)	archaic translations deleted
nescience	(poet., liter.)	deleted
ness	(arch.)	deleted
=ness	suff.	fragments deleted as a rule
nest		several example phrases and a compound deleted
nesting		example phrase deleted
nestle		
nestling		
Nestor	(Gr. myth.)	
Nestorius	(hist.)	deleted
net ¹		several compounds deleted
net ² , nett		
nether	(arch. or joc.)	some example phrases deleted
Netherlands		example phrase and derivative Netherlandish deleted
Netherlands Antilles	(geog.)	
nethermost		
netsuke	(Jap.)	
nett		
netter		deleted
netting		

Of the 82 *Pharos One* lemmas, 32 (or 39%) were removed completely, while 14 had parts like example phrases, compounds or derivatives deleted. As can be seen from the *Pharos One* labels, the lemmas that have been removed are mostly dated, rare, technical or British or American peculiarities.

Note the new *Concise* entry **neomycin**. The *Concise* contains at least 1 000 new words, mostly in the EA half, with some, through translation, giving rise to new Afrikaans entries as well. Many of the new English entries were sourced from recent English dictionaries, in particular the *South African Concise Oxford Dictionary*.

In the above section from *Pharos One*, a mistake had crept in: among the compounds of **net**¹ is **network**, given with translations and an example phrase. However, further on, **network** occurs again in its own fairly extensive entry, with a verb homograph, several more example phrases, and two derivatives, **networker** and **networking**. In the *Concise*, this mistake has been rectified: **network** is retained as a compound of **net**¹, but only with a cross-reference to the independent lemma **network**.

3.2 Further simplification of the structure

The target of this restructuring was compounds, especially in lemmas containing an extensive list of them. Compounds that manifested in the two ways described below were removed from their original list and made into independent lemmas. The first group identified were those compounds which had achieved 'critical mass' by having their own compounds or derived compounds, i.e. compounds closer related to the preceding compound than the original lemma, or by having a goodly collection of example phrases of their own. The second group of compounds to be moved out were those that resemble derivatives more than compounds, especially those in which the joined element cannot readily be conceived of as an independent word.

The lemma **nood** (*Pharos One*, p. 376; *Concise*, p. 394) provides examples of both these kinds of compounds. Compounds of the first group that were made into independent lemmas are **noodhulp** (with an example phrase and three of its own compounds), **noodlot** (with several example phrases, a compound and two derivatives of its own) and **noodsaak** (with a verb homograph and three derivatives, each with an example phrase), while **noodwendig** (with an adverb homograph and a derivative) of which the etymology as a compound has become rather opaque, provides an example of the second group of compounds identified above.

Several more examples of the restructuring of compound lists can be seen at **mee** (*Pharos One*, p. 343; *Concise*, p. 358) and **saam** (*Pharos One*, p. 494; *Concise*, p. 518). However, both these sets are intricate — as examples — because of the concurrence of compounds of the variant prefixes **mede-** and **same-**. Where compounds of **mee** or **saam** were made independent, the variant derived compounds with **mede-** or **same-** were added to the entry (with cross-referencing from their alphabetically correct place). See for example the *Concise* entries of **meedeel** and **meeding**, and of **saamdrom**, **saamflans**, **saamhang**, **saamkom**, **saamleef**, **saamloop**, **saamroep**, **saamsmelt**, **saamsnoer**, **saamstel**, **saamsweer**, **saamtrek**, **saamval**, **saamvat**, **saamvloei**, **saamvoeg** and **saamwerk**. Extensive

reference to the electronic archives was required to determine by means of frequency counts which of, for example, **mededinging** or **meedinging**, or **saamtrekking** or **saamtrekking**, should receive precedence.

4. Conclusion

When Pharos Dictionaries was established in 1996, its first priority was to develop a replacement for the TW, a reference work which had set the standard for bilingual AEEA dictionaries and was widely used, but was in danger of becoming badly outdated. In 2005 Pharos released the much anticipated *Pharos One*, whose status as a worthy successor to the TW was soon recognised. In 2006 it was honoured by both the Afrikaanse Taal- en Kultuurvereniging (with a Woordwys Prize) and the South African Translators' Institute (for Outstanding Translation).

The outstanding features of *Pharos One* are its user-friendliness and accessibility, its wealth of modern material and the sensitive way in which older material is treated. All its entries offer information about morphology, pronunciation, inflection and usage and the structure of the lemmas is open and logical. The new material places *Pharos One* among the most up-to-date South African reference works and enables a confident interchange between modern English and Afrikaans. The collection of older material in *Pharos One* is comprehensive and extensively labelled, particularly regarding words and expressions which may have become offensive.

Despite its enhanced features, *Pharos One* has yet to equal the overall sales rate of TW8. This could be due to the size and price of *Pharos One*, but the significantly reduced role of Afrikaans in the present-day South Africa must have a decisive impact too. To counter the first-mentioned factors Pharos Dictionaries decided to produce a scaled-down version of *Pharos One*. This one-third reduction became the *Concise*, achieved by culling from *Pharos One* dated material, overly technical vocabulary and words and expressions which are particularly British, American or Australian. With dimensions of 18 cm x 25 cm x 5 cm and a mass of 1,8 kg, the *Concise* is considerably smaller and more affordable than *Pharos One*.

The average monthly sales of the *Concise*, released in 2007, are steady and match those of *Pharos One*. Significantly, this means that the combined average monthly sales of *Pharos One* and the *Concise* are now on par with those of TW8 over its entire 20-year print run. In this respect the production of the *Concise* has been justified: not only does it offer the user the best of *Pharos One* at a reduced price, its presence in the market has augmented the sales rate of Pharos's comprehensive AEEA dictionaries to a level comparable to that of TW8. This is no mean feat, considering that the TW benefited hugely during the time in which Afrikaans was one of only two official languages.

These developments demonstrate not only that the work of Pharos's editors is held in high esteem by their peers in cultural and language bodies, but

also that the general dictionary-buying public continues to recognise the enduring quality of dictionaries bearing the Pharos brand.

References

The name(s) of the editor(s)/compiler(s) is (are) included if stated on the title page. Unless it is the first edition, the edition number is given only if it is specifically mentioned.

- Bosman, D.B., I.W. van der Merwe, and L.W. Hiemstra (Original Eds.); Joubert, P.A. and J.J. Spies (Eds.: Eighth Edition).** 1984⁸. *Tweetalige Woordeboek/Bilingual Dictionary*. Cape Town: Tafelberg.
- Collins English Dictionary.* 1998⁴. Glasgow: HarperCollins Publishers.
- Du Plessis, M.** 1999. *New Words/Nuwe Woorde English–Afrikaans*. Cape Town: Pharos.
- Eksteen, L.C. (Ed.).** 1997¹⁴. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans*. Cape Town: Pharos.
- Gove, P.B. (Ed.).** 1961. *Webster’s Third New International Dictionary*. London: G. Bell & Sons.
- Hanks, P. (Ed.).** 1985². *Collins Dictionary of the English Language*. London/Glasgow: Collins.
- Hockey, P.A.R., W.R.J. Dean and P.G. Ryan (Eds.).** 2005⁷. *Roberts Birds of Southern Africa*. Cape Town: John Voelcker Bird Book Fund.
- <http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>
- <http://www.merriam-webster.com>
- Joubert, P.A.** 1997. *Tweetalige Frasewoordeboek/Bilingual Phrase Dictionary*. Cape Town: Pharos.
- Labuschagne, F.J. and L.C. Eksteen (Eds.).** 1993⁸. *Verklarende Afrikaanse Woordeboek*. Pretoria: J.L. van Schaik.
- Martin, W. and G.A.J. Tops (Eds.).** 1984. *Van Dale Groot Woordenboek Engels–Nederlands*. Utrecht/Antwerpen: Van Dale Lexicografie.
- Odendal, F.F. and R.H. Gouws (Eds.).** 2000⁴. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Johannesburg: Perskor.
- Pearsall, J. (Ed.).** 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Pearsall, J. (Ed.).** 1999¹⁰. *Concise Oxford Dictionary*. Oxford: Oxford University Press.
- Pharos Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary.* 2005. Cape Town: Pharos.
- Pharos Afrikaans–Engels/English–Afrikaans Kernwoordeboek/Concise Dictionary.* 2007. Cape Town: Pharos.
- Schoonees, P.C., F.J. Snijman, D.C. Hauptfleisch, D.J. van Schalkwyk and W.F. Botha (Eds.).** 1950–2005. *Woordeboek van die Afrikaanse Taal Deel I–XII A–Q*. Pretoria: Government Printer/Stellenbosch: Bureau of the WAT.
- Springer, O. (Ed.).** 1962. *Langenscheidt’s Encyclopaedic Dictionary Part 1 English–German*. Berlin: Langenscheidt.
- The American Heritage College Dictionary.* 1993³. Boston/New York: Houghton Mifflin Company.
- The Dictionary Unit for South African English.** 2002. *South African Concise Oxford Dictionary*. Cape Town: Oxford University Press Southern Africa.

Wine and Words: A Trilingual Wine Dictionary for South Africa*

Michele F. van der Merwe, *Faculty of Education and Social Sciences, Cape Peninsula University of Technology, Wellington Campus, Wellington, Republic of South Africa (vandermerwem@cput.ac.za)*

Abstract: The South African wine industry identified the need for a special-field on-line dictionary on viticulture and oenology in Afrikaans, English and isiXhosa. The dictionary provides information on wine terminology as well as linguistic information on the use of such terminology. The purpose of this article is to give a description of the project. The process of compiling the dictionary is described, from the co-operation between the wine industry and lexicographers to the intended target users and the choice of languages of the dictionary. Functions of the dictionary are discussed, with reference to specific user situations, namely text production, text reception and translation. A system of labels has been designed for the dictionary and its benefit for the user is explained. In assisting the user to make an informed choice of a term, the notion of proscriptiveness has been followed in the presentation of information in the wine dictionary.

Keywords: TRILINGUAL WINE DICTIONARY, SPECIALISED LEXICOGRAPHY, VITICULTURE AND OENOLOGY TERMS, ON-LINE DICTIONARY, TARGET USERS, USER SITUATIONS, FUNCTIONS, TEXT RECEPTION, TEXT PRODUCTION, TRANSLATION, LABELS, ENCYCLOPEDIAIC KNOWLEDGE, LINGUISTIC KNOWLEDGE, PROSCRIPTION

Opsomming: Wyn en woorde: 'n Drietalinge Wynwoordeboek vir Suid-Afrika. Die Suid-Afrikaanse wynbedryf het die behoefte aan 'n aanlynwoordeboek oor wyn- en wingerdkunde in Afrikaans, Engels en isiXhosa geïdentifiseer. Die woordeboek verskaf inligting oor wynterminologie, sowel as taalkundige inligting oor die gebruik van sulke terminologie. Die doel van hierdie artikel is om 'n beskrywing van die projek te gee. Die samestellingsproses van die woordeboek word beskryf, vanaf die samewerking tussen die wynbedryf en die leksikograwe, tot die voorgestelde teikengebruikers en die keuse van die tale van die woordeboek. Funksies van die woordeboek word bespreek, met verwysing na spesifieke gebruikersituasies, naamlik teksproduksie, teksresepse en vertaling. 'n Etiketstelsel is vir die woordeboek ontwerp en die voordeel daarvan vir die gebruiker word verduidelik. Ten einde die gebruiker te help om 'n ingeligte keuse van 'n term te maak, is die begrip van proskripsie gevolg by die aanbieding van die inligting in die woordeboek.

Sleutelwoorde: DRIETALIGE WYNWOORDEBOEK, GESPECIALISEERDE LEKSIKOGRA-

* This article was presented as a paper at the Twelfth International Conference of the African Association for Lexicography, organised by the Tshwane University of Technology in collaboration with the IsiNdebele National Lexicography Unit, at the Soshanguve Campus, Pretoria, Republic of South Africa, 27–29 June 2007.

FIE, WINGERD- EN WYNKUNDETERME, AANLYNWOORDEBOEK, TEIKENGEBRUIKERS, GEBRUIKERSITUASIES, FUNKSIES, TEKSRESEPSIE, TEKSPRODUKSIE, VERTALING, ETIKETTE, ENSIKLOPEDIËSE INLIGTING, TAALKUNDIGE INLIGTING, PROSKRIPSIE

1. Why a dictionary of wine terms?

The need for a special-field dictionary of wine terms was identified by the wine industry in 2005. Already in 1973, a South African *Wine Dictionary* was compiled by the Language Services Bureau with Afrikaans and English as language pairs. Since this dictionary is merely a list of terms in the source language, accompanied by their target-language equivalents, it is of little assistance to the present-day user. In this *Wine Dictionary*, no explanations of and no syntactic information on terms are provided. Many of the terms had to be replaced by new ones as some of them were no longer in use in the wine industry, for example, *keg*. The rapid development in the field of wine biotechnology necessitated the creation of many new terms to account for new inventions in the field, for example, *biofilm*, *biogenesis*, *bioinformatics* and *biosensor*. Being dated, the *Wine Dictionary* could not make provision for new terms to describe developments in any field relating to the wine industry for the last 35 years. Providing neither semantic nor grammatical information, such a dictionary at best confirms the expert's own assumptions, but laypeople and semi-experts alike will find it unsuitable for text reception and translation purposes. The 1973 dictionary, containing only a restricted number of terms, is not widely known among wine industry members, and, as it is out of print, it was decided to compile a new wine dictionary.

The *South African Trilingual Wine Dictionary* (SATWD) makes provision for data on viticulture and oenology. The subject field viticulture includes organic cultivation and production, cultivation, soil science, plant biotechnology, vine viruses, plant protection and plant improvement. The subject field oenology includes production technology, bottling, packaging and distribution, by-product handling and waste product handling, brandy and distilling, and microbiology. In short, the SATWD transfers information on the physiology of the vine, the fruit of the vine as well as the wine-making process.

It was decided to use Trados Multiterm to create an electronic dictionary with the advantage of being cost-effective to the industry. On completion the database would be available on-line to the wine industry and to the general public. Access would be via two websites, namely the Sawis Library website and the Winetech website.

Another advantage of an on-line dictionary would be to manage data effectively, by updating it on a continuous basis, adding or deleting terms as required. It could prove to be very effective, especially in a field like the wine industry where there is so much development with new concepts gaining currency and subsequently new terms being created on a regular basis.

After the launch of the SATWD on-line, the compilers will hopefully

receive continuous feedback from the industry. Users, be they laypeople or subject-field specialists, could contribute to the dynamism of the dictionary by continuous communication on-line. The advantage of user participation would be the input on translations in Afrikaans and isiXhosa, as well as definitions of terms. Provided that sufficient comments and feedback are received, it could lead to the improvement of the dictionary by reflecting the views of users and their use of terms and translations.

The purpose of the article is to describe the project by sketching the initialisation of the project, describing the purpose and function of the dictionary, and referring to other special-field dictionaries. The extensive labelling system and proscriptive treatment of terms are also discussed.

2. Co-operation between the wine industry and lexicographers

Winetech, coordinator of research, training and technology in the wine industry, and Sawis, responsible for the collection, processing and dissemination of industry information, jointly fund the project. This is a first for the wine industry to be engaged in a lexicographic endeavour. There are examples of agricultural industry boards involved in the compiling of special-field dictionaries in South Africa. In 1958 *Vertalende en Verklarende Woordeboek vir die Wolbedryf* (ed. A.J. Hanekom) was published by the South African Wool Board and in 1990 *Directory of Cotton Terms for South Africa* (ed. K.W. Sanderson) was produced by the South African Cotton Board.

The Preface to the *Encyclopedic Dictionary of Gene Technology* (Bergenholtz and Kaufmann 1998: 5) rightly states: 'No technically related dictionary article shall be designed solely by a linguistic nor by a technical expert. Only the interdisciplinary co-operation can produce a reference work which can be both lexicographically and technically satisfactory.' The SATWD is a practical example of such co-operation and consultation. Lexicographic planning was done by a lexicographer, resulting in a dictionary plan which was then submitted to subject-field specialists, as well as industry bodies. Input was received from various subject-field specialists, mostly academics and the industry body Winetech. During the compilation of the SATWD there was continuous consultation between the lexicographer, information specialist and subject-field specialists. Consultation related to issues such as the inclusion of terms, the translation of terms, definitions of terms and the use of different subject labels. Since the SATWD covers such a wide range of subject fields within the field of viticulture and oenology, it was essential to consult with experts on these various fields.

3. Choosing the languages

The wine industry in South Africa consists of three geographical units, namely the Western Cape, the Northern Cape and Kwazulu-Natal. Since most of the

wine-making activities are based in the Western and the Northern Cape, it was decided to focus on the official languages of these provinces, namely Afrikaans, isiXhosa and English. The SATWD seeks to fulfil the language needs of users of the designated language groups regarding wine terminology in South Africa. Many Afrikaans-speaking farmers are involved in the wine industry. However, there is a growing need also for isiXhosa speakers to join the industry. With more emphasis on the training and education of emerging farmers the development of isiXhosa wine terminology is therefore essential.

The pedagogical aspect of the dictionary is important for it can be a useful tool in the training of Afrikaans, English and isiXhosa students at tertiary level. There is a need for students to be able to use the technical terminology of the wine industry, be that at university or agricultural college level. The SATWD can serve as an important communication tool in the training of especially isiXhosa students, since a large number of words were translated into isiXhosa during the compilation of the SATWD. This is the first time that isiXhosa would feature as a member of a language pair in a specialised dictionary of wine terminology.

It is assumed that mother-tongue education could play a major role in providing access to science and technology. Such a multilingual specialised dictionary could serve as an important learning aid during tertiary education, be that for lecturers to compile course material for students or for students to look up translations and definitions of terms. Functions of the dictionary to be used in these instances would be those of text production, translation and text reception.

Another target group of users to benefit from the pedagogical aspect of the SATWD would be people belonging to wine circles wishing to be trained in organoleptic evaluations of wine. According to Brochet and Dubourdieu (2001: 187), wine is one of the more often described foods, wine literature being widespread with numerous journals, books and reviews presented to the public and to experts or professionals. Those who fail to describe wine flavours, unable to speak about its taste in a professional manner, are usually considered as 'not knowing anything about wine.' Winemakers, professional critics, oenologists, and amateurs have developed a language or vocabulary which describes the sensory properties of wine. This is used to exchange information on sensory data and to analyze this information. The SATWD contains a large number of terms for the organoleptic evaluation of wine, and especially in the case of Afrikaans and isiXhosa, the vocabulary of these languages has definitely been extended. The Afrikaans wine magazine *Fynproe* has also contributed greatly to the extension of the Afrikaans wine-tasting vocabulary.

Being an on-line dictionary, the SATWD can contribute indefinitely to the extension of languages, with the possibility of including other indigenous languages, like isiZulu (Kwazulu-Natal forms another geographical unit of the wine industry) as well as the languages of the traditional European wine-producing countries like France, Spain and Germany.

4. Identifying the target users

In August 2005, a dictionary plan was drawn up and target-users were identified. Bergeholtz and Kaufmann (1997: 101) distinguish three groups of target-users for LSP dictionaries: laypeople, semi-experts, and experts. Most of the target-users of the SATWD could be described as semi-experts, having a higher level of knowledge than laypeople. They include viticulturists, vintners, wine-farmers, wine-lovers, wine-writers, wine-marketers, wine-consultants, wine and viticulture lecturers and students, extension officers and information officers, as well as translators. As this user-group can be described as heterogeneous, provision needs to be made for covering problems relating to the subject field in language comprehensible to laypeople and semi-experts alike.

Excellent provision was made for such a heterogeneous group in the *Encyclopedic Dictionary of Gene Technology* (EDGT) (Kaufmann and Bergeholtz 1998: 8) where the compilers opted for an encyclopaedic dictionary instead of the traditional reference work providing encyclopaedic data or conveying linguistic data. In the case of one or more subject fields, the former are called specialised encyclopaedias and the latter specialised dictionaries. This dictionary provides factual data and specialised linguistic data on gene technology. The aim of the reference work is to be an 'aid to understand purely encyclopaedic relationships and an aid in the communications of topics on gene technology'. One of the aids implemented in this dictionary is a very long introduction, called the encyclopaedic section on the subject of gene technology. It forms the subject-field component of the dictionary, giving an overview of the systematics forming the basis of gene technology. This overview can be read as a small work in its own right, completely independently from the dictionary, to obtain basic information on the technical basis of molecular biology and gene technology. In a dictionary context, it is intended as an aid for those readers requiring a concise account.

A similar subject-field component should be compiled in Afrikaans, English and isiXhosa for the SATWD. Such an introduction can feature data on viticulture, for example the physiology of the vine, and data on oenology, for example the wine-making process, as well as the appreciation of wine. Users would have access to data on the different stages of the wine-making process. They could read it as a background or they could use it as reference when looking up specific information regarding a term or a definition encountered in the dictionary. The subject-field component would then serve as an integrated text in the dictionary which the user can consult when, having looked up the meaning of a term in the dictionary, the need arises to gain more information on the specific term.

5. Functions of the SATWD

As lexicographers become more aware of the notion of the user-perspective

and try to employ user-orientated techniques in dictionaries, the typology of traditional dictionaries change. It is therefore not strange to find more dictionaries with explanatory as well as translational qualities. Change has also occurred in special-field dictionaries, of which the *Encyclopedic Dictionary of Gene Technology* (Kaufmann and Bergenholtz 1998) is a good example.

Tarp (2005: 10) describes the EDGT as a two-way, bidirectional and multifunctional dictionary. Eight main user situations for which the dictionary is designed are specified. These are:

- (a) the reception of Spanish texts,
- (b) the production of Spanish texts,
- (c) the reception of English texts,
- (d) the production of English texts,
- (e) the translation from Spanish into English,
- (f) the translation from English into Spanish,
- (g) the information on subjects within molecular biology, and
- (h) the indirect text-related information about Spanish and/or English molecular biology language usage.

The functional theory (Tarp 2005: 7) concentrates on the extra-lexicographic user situations of which two main types can be distinguished: the communication-related and cognitive-related. The first is related to communicative situations such as L1 and L2 text production, L1 and L2 text reception, as well as L1→L2 translation and L2→L1 translation. The second is related to non-communicative situations where users want or need a certain kind of knowledge, for example linguistic, general encyclopaedic or specialised knowledge.

An example of two entries in the EDGT is provided. A user could look up the term *gene pool* in the English section and be given the translation in Spanish, as well as a definition of the term *gene pool*. The user would be able to use the dictionary for translation, text production in English and text reception in English. In the case of *gene preservation*, the Spanish translation is provided and the user would be able to use the dictionary for translation.

gene pool pool m genico

A gene pool is the total number of genes present in a population at a certain time. The addition of the second X-chromosome gene must have occurred sometime after the separation of South America and Africa, and hence of the gene pools of the New and old World monkeys, some 40 million years ago.

gene preservation preservacion f de genes

It is difficult to typologise the SATWD according to traditional dictionary categories. It can at best be described as a multifunctional dictionary, having explanatory as well as translational qualities. It is aimed at text reception, text production and text translation. A translation of a term is supplied in two languages, e.g. Afrikaans and isiXhosa.

The purpose of the SATWD would be firstly to supply data on the subject

fields viticulture and oenology, as well as linguistic information on terms.

The explanatory part of the dictionary includes a definition of the term in English, a subject label and an indication of the word class/part of speech. Word class of terms is provided, since that is the only syntactical information presented to the user as no illustrative material is given, as shown by the following two examples from the SATWD.

abscisic acid *Part of Speech* n *Definition* a hormone that occurs naturally in vines and regulates growth, fruit-ripening and physiology. (CD) *Subject* Grapevine|Hormone
Afrikaans **absiensuur**
Xhosa **i-asidi i-abhisisikhi**

gene silencing *Part of Speech* n *Source* Wynboer Technical Yearbook 2006/7, p 12.
Definition preventing or suppressing genes being translated into proteins. (CD)
Subject Biotechnology
Afrikaans **geendemping**
Xhosa **ukucinywa komyalezo wofuzo**

The SATWD is designed for the following user situations:

- (a) text reception in English,
- (b) text production in English,
- (c) translation from English into Afrikaans, and
- (d) translation from English into isiXhosa.

A list of translational equivalents for Afrikaans and isiXhosa terms will be provided and the following user situations can occur:

- (e) translation from Afrikaans into English,
- (f) translation from Afrikaans into isiXhosa,
- (g) translation from isiXhosa into Afrikaans, and
- (h) translation from isiXhosa into English.

In the South African context, there is a resemblance between the SATWD and *A Glossary of Soil Science*, published in 1990 (second edition 1995) by the Soil Science Society of South Africa. In this specialised dictionary, translational equivalents and definitions are presented, with the dictionary fulfilling a text reception as well as a translational role. The difference between the two dictionaries is that the SATWD provides explanatory sections only in English, leaving the user with limited lexicographic treatment of a term in isiXhosa and Afrikaans. This is not the ideal situation, also from a pedagogical point of view where mother-tongue knowledge transfer is more advantageous to the user. Hopefully it would be possible to include explanatory sections in Afrikaans and isiXhosa at a later stage of the project in order to ensure text reception in these languages as well.

In the near future, the wine industry plans the Foresight Project and the SATWD could form part of the information system provided. The purpose of the Foresight Project (Spies 2007) is

establishing an interactive IT-based information platform at SAWIS to develop a 'Vinipedia' in line with the highly successful global Wikipedia interactive internet service. This would be a learning information search design where the best minds make contributions in specific subject areas. It evolves and improves with the contributions of professionals — becoming ever more legitimate as a search engine for initial inquiries. Such a South African 'Vinipedia' could form the basis for a strategic overview service that is always up to date — professional and legitimate.

If the SATWD becomes part of the Foresight Project, more possibilities for text reception could open to users. The outer texts in the Foresight Project could be consulted to gain information on terms used in the dictionary, for example more information on viticulture or oenology. It could serve as an extension of the dictionary, supplying more data to the user to enable text reception. An example of text reception would be a user collecting data on the topic of vine diseases or wine tasting terms.

6. Labels, from vine to wine

In compiling the SATWD, the presumed needs of the different user-groups were kept in mind and a handy device, the use of subject labels, was designed to assist users to ensure maximum transfer of information. According to Gouws and Prinsloo (2005: 129), lexicographic labels are frequently used in the comment of semantics to give explicit contextual guidance. They distinguish between subject field labels, stylistic labels and chronolectic labels.

In the case of the SATWD, a whole set of subject-field labels was chosen to assist the user in the transfer of semantic information. As the majority of target-users could be viewed as semi-experts, they would need considerable guidance to understand terms fully and completely. A large variety of subject-field labels are used to indicate usage within a semantic network.

During the planning stage of the dictionary, a hyponymous relationship was created between labels, for example *viticulture: grapevine disease*. A term like *root plague* is placed in a broader semantic field, namely *viticulture*. The field was then narrowed down by stating that the term forms part of a subdivision *grapevine disease*. After consultation with subject-field experts, the labelling system was slightly changed.

It is assumed that a term would either belong to one of the two subfields, viticulture and oenology to be lemmatised in the SATWD. According to subject-field specialists, it would be a repetition to state, for example, that *grapevine disease* falls under *viticulture*, thus labelling *root plague* with *viticulture: grapevine disease*. Instead it was opted to use only the label *grapevine disease*.

Since the scope of the subject fields viticulture and oenology are so vast, they have been divided into smaller fields more readily to make transfer of information. Under the field of viticulture, one would find subfields such as *cultivar*, *cultivation practice*, *equipment*, *fertilization*, *grapevine deficiency*, *grapevine*

disease, grapevine morphology, grapevine pest, grapevine physiology and growth regulator, while under the field of oenology, one would find subfields such as *barrel size, bottle size, cooperage, distillation, varietal wine, wine aroma wheel, wine fault, wine-making, wine region and wine style*. If a term belongs to more than one subfield, both of the subfield labels were used to indicate such a relation, as in the case of *2, 4, 6-trichloroanisol* where a relation to both *wine fault* and *wine making* is shown.

2,4,6-trichloroanisol *Part of Speech* n *Abbreviation* TCA *Definition* a chemical compound which is responsible for the wine being corked (wine fault). (CD) *Subject* Wine fault/Wine making
Afrikaans **2,4,6-trichloranisool**
Xhosa **itrirhloranisoli-2,4,6**

In the SATWD, terms relating to subject fields other than viticulture and oenology have also been included. The criteria applied were that whenever a term is used in viticulture or oenology, it would be recorded in the dictionary. Subject fields such as *biochemistry, biotechnology, botany, chemistry, genetics, law, marketing, microbiology, plant pathology and soil science* were added. Terms in these related fields have been labelled accordingly, ensuring successful information transfer. If there is overlapping between different subject fields where a term belongs to both fields, it is indicated with the use of the appropriate subject labels. In the case of *bud* three subject labels were used to show a relation to *botany, grapevine anatomy and viticulture*.

accessory bud *Part of Speech* n *Definition* either the tertiary or secondary bud as part of the compound bud. (CD) *Subject* Botany|Grapevine anatomy|Viticulture
Afrikaans **newe-oog**
Xhosa **imehlo**

As many terms as possible were labelled in order to help the user with sufficient information transfer, trying to present a bigger semantic picture for the user. To know under which subject field or subfield a term falls, could be of great benefit to a user trying to understand a specific term. A broader semantic insight could subsequently help the user with better text reception and text production.

7. Proscriptiveness

The aim of the SATWD is to provide information on the subject fields of viticulture and oenology, as well as linguistic information on terms. During the compilation of the dictionary, one of the challenges posed to the lexicographer was how to deal with the notion of prescriptiveness versus descriptiveness. In a dictionary with a prescriptive point of view, the lexicographer would present the only 'correct' term or translation according to his/her research or findings and a specific linguistic variant can be prohibited. In a dictionary with a

descriptive point of view, the lexicographer would give all possible terms and translations according to his/her research and findings and it would be left to the user to choose which term or variant to use.

In the case of the SATWD, there is a need to leave it to the user to choose from the given terms, but the dictionary must provide him/her with sufficient information to make an appropriate choice. The lexicographer should ask what would benefit the user most. If the dictionary is to be a practical instrument for the user, it would have to fulfil the function of container of knowledge — knowledge of the world and of language. There is a continuous tension between the use of some terms by subject experts and wine industry bodies and the way ordinary users perceive meanings and use words. In order to deal with this, the lexicographer can fulfil the role of a mediator by assisting users in the choice of an appropriate term. In the SATWD, a term would be presented with an editing note indicating whether it is the term preferred by the wine industry or not. The lexicographer could recommend using another term or provide another term reflecting the real language usage of ordinary language speakers. This method could be seen as 'proscriptive'. According to Bergenholtz (2003: 77), proscription entails the presentation of information on language use and assistance with the choice of the correct variant from more than one possibility: 'In practice proscription would be a selective description. Unlike description where every single case is invented individually and described on its own, proscription throughout compares similar cases systematically and makes the same decisions in each case.'

The wine industry has imposed restrictions on the use of certain terms. In the 1973 *Wine Dictionary*, the term *keg* was used. Since then it was superseded by the word *barrel* which the wine industry prefers to use instead of *cask*, although linguistically there is no difference between the two terms. Another example of such a restriction is the word *variety* which is not used by the wine industry, the preference being for *cultivar*.

In many cases, ordinary wine drinkers are unaware of restrictions imposed on certain terms by the wine industry. In the case of *boxed wine*, the wine industry refers to a specific packaging method, translated as *tapvat* in Afrikaans. However, in many cases ordinary users would refer to it as *bokswyn*. There is a difference in register between the two terms, with *bokswyn* being the informal term. Ordinary wine drinkers also attach the meaning of a lesser quality wine to the term *bokswyn*. Naturally the wine industry would be very aware of the notion of quality as demonstrated by the use of certain terms. In the case of *bulk wine*, the wine industry prefers the use of *grootmaat* instead of *stortmaat*, which has a connotation of lesser quality. The term *drinking wine* (*drinkwyn* in Afrikaans) which was used in the wine industry in official documents during 2002–2005, has been replaced by *wine*, but there are still ordinary wine drinkers referring to wine in this way. In marketing circles, it is fashionable to use the term *handcrafted wine* on wine labels, but according to the wine industry no such term officially exists. It would then be the duty of the lexico-

grapher to ensure sound transfer of information to the user and, since the term occurs in texts, account for it in the dictionary.

Owing to the specific political environment in South Africa, sensitivity has developed around the use of certain words. In the case of the wine industry, one of these words is *dopstelsel* which describes the infamous way in which some farm labourers on wine farms were remunerated years ago. Although the wine industry is sensitive to the word because of its political connotations and the feelings it can arouse, the word was included in the SATWD. As it is referred to in many documents of the wine industry, it would be to the benefit of the user that the word is defined.

dop system *Part of Speech* n *Definition* the illegal system of giving inferior wine to Cape coloured vineyard and cellar workers in partial lieu of wages. (CD)
Afrikaans **dopstelsel**
Xhosa **uhlawulo ngotywala**

It was decided, however, not to include *boys' wine* (*volkswyn* in Afrikaans), referring to an inferior wine, since the word is no longer in use. Because it is not based on historical principles, the SATWD does not include historical words.

Another restriction on terms in the wine industry would be one imposed on South Africa by the European Union. According to the EU, the word *champagne*, for example, may only occur on a wine etiquette if the content of the bottle was made according to the Champagne method used in France. In South Africa, a similar method is used, but it is called 'Cap Classique'. Champagne made according to the Cap Classique method is more expensive. One would therefore rather find more sparkling wine on the South African market. Sparkling wine is not made according to the Cap Classique method and is not of the same quality. Many users would not know the difference between *champagne* and *sparkling wine*, incorrectly using the term *champagne* for *sparkling wine*. In this case, the lexicographer could present information proscriptively by including the term *champagne* in the dictionary and providing it with an editorial note, explaining the situation to the user, but also recognising the application of the word by ordinary users.

8. Conclusion

The SATWD was planned with a specific user group in mind who could consult the dictionary in different user situations for different functions. One of the advantages is the on-line accessibility of the dictionary. It is the vision of the compilers of the dictionary that it should become a practical, handy tool for the user, be that person a member of the wine industry, or just an ordinary wine-drinker.

References

Dictionaries

- Kaufmann, U. and H. Bergenholtz (Eds.).** 1998. *Encyclopedic Dictionary of Gene Technology. Volume 1: English (with Spanish Equivalents). Volume 2: Spanish (with English Equivalents)*. Toronto: Lugus Libros.
- Hanekom, A.J. (Ed.).** 1958. *Vertalende en verklarende woordeboek vir die wolbedryf. Engels-Afrikaans. Met verklarings in Afrikaans*. Pretoria: South African Wool Board.
- Sanderson, K.W. (Ed.).** 1990. *Directory of Cotton Terms for Southern Africa/Handleiding van katoenterme vir Suidelike Afrika*. Pretoria: South African Cotton Board.
- Terminology Section, Language Services Bureau in Collaboration with the Oenological and Viticultural Research Institute.** 1973. *Wine Dictionary (including Viticulture, Winemaking, Wine Chemistry)/Wynwoordeboek (insluitende wingerdbou, wynbereiding, wynkunde)*. Pretoria: Department of National Education.
- Van der Watt, H.v.H. and T.H. van Rooyen.** 1990. *A Glossary of Soil Science*. Pretoria: The Soil Science Society of South Africa.

Other sources

- Bergenholtz, H. and U. Kaufmann.** 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. Tarp, S. (Ed.). 1997. Specialised Lexicography and Terminology. *Hermes, Journal of Linguistics* 18: 91-125.
- Bergenholtz, H.** 2003. User-orientated Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Brochet, F. and D. Dubourdieu.** 2001. Wine Descriptive Language Supports Cognitive Specificity of Chemical Senses. *Brain and Language* 77: 187-196.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Spies, P.H.** 2007. *South African Wine Industry Business Foresight Project: Framework for the Design of a Wine Industry Foresight and Business Intelligence Service*. Available in the SAWIS Industry Library.
- Tarp, S.** 2005. The Pedagogical Dimension of the Well-conceived Specialised Dictionary. *Iberica* 10: 7-21.

A Survey of the Dictionary Use of Gabonese Students at Two South African Universities*

Guy-Modeste Ekwa Ebanéga (*guymodeste_e@yahoo.fr*)
and

Fatima Tomba Moussavou (*tombafatima@yahoo.fr*),
*Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch,
Republic of South Africa*

Abstract: This article presents a preliminary study on Gabonese users' knowledge, opinions, attitudes and habits of the use of dictionaries, and their dictionary culture. It is based on the principle that the lexicographer should know the target users and their needs (Householder and Saporta 1962: 279). The aim of the article is to present and discuss research findings of a lexicographic survey conducted among 100 Gabonese students at the University of Stellenbosch and the Cape Peninsula University of Technology with regard to dictionary use and culture. In this article, we discuss the objective of the research, the research subjects, the research methods, the research coverage, the research findings, the description and the critical analysis of the data.

Keywords: SURVEY, QUESTIONNAIRE, GABONESE USERS, DICTIONARY CULTURE, NEEDS

Résumé: Une enquête de l'utilisation du dictionnaire des étudiants gabonais à deux universités sud-africaines. Cet article présente une étude préliminaire sur la connaissance des utilisateurs gabonais, leurs avis, leurs attitudes et leurs habitudes de l'utilisation des dictionnaires, et leur culture de dictionnaire. Il est basé sur le principe selon lequel le lexicographe devrait connaître les usagers cibles et leurs besoins (Householder et Saporta 1962: 279). Le but de l'article est de présenter et discuter des résultats de recherches d'une enquête lexicographique conduite parmi 100 étudiants gabonais à l'Université de Stellenbosch et à l'Université Cape Peninsula de Technologie en ce qui concerne l'utilisation et la culture du dictionnaire. Dans cet article nous discutons l'objectif de la recherche, les informants, les méthodes de la recherche, l'assurance de la recherche, les résultats de la recherche, la description et l'analyse critique des données.

Mots-clés: ENQUÊTE, QUESTIONNAIRE, USAGERS GABONNAIS, CULTURE DU DICTIONNAIRE, BESOINS

* This article is an adapted version of a paper presented at the Eleventh International Conference of the African Association for Lexicography, organized by the Tshivenda National Lexicography Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5–7 July 2006.

Introduction

Lexicographically speaking, three periods characterize Gabonese lexicography. The first is the missionary period, followed by the colonial period and eventually the modern period. The first and the second are the periods of the missionaries and the colonial administrators. They are the pioneers in the compilation of dictionaries in the Gabonese languages, which are all bilingual (or translation) dictionaries biased towards French and aimed at foreign users, the French speakers. The third is the modern period when national lexicographers trained in metalexicographical principles started compiling dictionaries aimed at the Gabonese users.

This article describes and discusses a preliminary survey on Gabonese users' knowledge of, opinions about, attitudes towards and habits regarding the use of dictionaries, and their dictionary culture. Most studies of dictionary use have therefore relied on the retrospection of users by means of questionnaires or interviews (for example Tomasczyk 1979, Béjoint 1981, Bogaards 1988, and Atkins and Varantola 1998).

This article is based on the premise that lexicographers should know the target users and their needs (Householder and Saporta 1962: 279).

The aim of this article is to present and discuss research findings of a lexicographic survey conducted among 100 Gabonese students at the University of Stellenbosch and the Cape Peninsula University of Technology with regard to dictionary use, their dictionary culture.

1. Problem statement

Hartmann (1989), Gouws (1989) and Wiegand (1998) emphasize an approach of compiling dictionaries guided by the user-perspective. The user-perspective, so prevalent in modern-day metalexicography, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. Hartmann (1989) aptly indicates that the compilation of any new dictionary has to be preceded by an in-depth needs analysis. This analysis can only be done once the intended target users have been identified unambiguously. It should reflect not only the needs but also the reference skills of the target users. An analysis of users' needs should precede dictionary design. According to Gouws and Prinsloo (2005: 39), good dictionaries do not only display a linguistically sound treatment of a specific selection of items. They are products that can be used as linguistic instruments by their respective target user groups.

According to Tarp (2002: 67), the starting point when planning, compiling or reviewing a dictionary should always be the users, the user situations and the complex of problems related to these situations which could finally cause the user to consult a dictionary for assistance. Householder and Saporta (1962: 279) emphasize that 'dictionaries should be designed with a special set of users

in mind and their specific needs'. This stresses the importance of understanding the needs of the dictionary users before compiling a dictionary.

For De Schryver and Prinsloo (2000: 1-31), the feedback from the users should be obtained while the compilation of the dictionary is still in progress, a process referred to as 'simultaneous feedback'. Atkins and Varantola (1998) and Bogaards (1998) are good examples of the serious efforts currently being made to research dictionary use. The problem however is that such research is generally carried out at a stage when dictionaries are already completed and published. Feedback from target users can therefore only be implemented in forthcoming editions of these dictionaries.

Up to now, no survey or empirical research has been done to establish the profile of Gabonese people as dictionary users. The statement of Whitcut (1986: 111) 'we know who we are, but who are they?' is still valid today and particularly in the Gabonese environment.

2. Objective of the research

The main purpose of this research, which was carried out during the academic year 2005–2006, was to draw a profile of Gabonese students as dictionary users. Indeed, the intention was to ascertain their knowledge about, opinions on, and habits in the use of dictionaries. The general goals were the following:

- To gather information about the dictionary users (the frequency of their use of dictionaries, their difficulties with the lookup process and the causes of these difficulties, the instruction they received to acquire dictionary skills, etc.)
- To find out the users' attitude towards different reference books (general dictionaries, bilingual dictionaries, encyclopaedic dictionaries, etc.)
- To know with which type of dictionary to start if dictionaries are to be compiled in the Gabonese languages.

3. Research subjects

The survey was carried out among 100 undergraduate and postgraduate Gabonese students, who are studying at the University of Stellenbosch and the Cape Peninsula University of Technology. These students came to study in the Western Cape under the convention between Gabonese universities and South African universities. Many Gabonese also study abroad, particularly in France, at university and graduate levels.

Because of money constraints, the investigation was undertaken only among Gabonese students in South Africa. The behaviour of Gabonese dictionary users studying in Gabon, for example at the Omar Bongo University, Libreville, and the University of Science and Technology of Masuku, near

Franceville, could therefore not be examined. For the purpose of this survey, the students were asked some personal information such as their gender, their age, their course of study, their degree, and their study subjects. Enquiries were also made about the language they mainly speak at home, the foreign languages they have studied and their ability to write in their home language. The participants in this survey were 36 females and 64 males with ages varying between 19 and 35 years.

- Most Gabonese students have different mother-tongue backgrounds. The distribution of the mother-tongue speakers is as follows:

Fang: 29	Lembaama: 7	Myene: 4	Tsogho: 1
French: 26	Yinzebi: 5	Lekanighi: 2	Teke: 2
Yipunu: 5	Awandji: 1	Mbebe: 1	Lumbu: 1
Mahongwe: 1	Bemba: 1	Ghisira: 1	Aduma: 3
Ndumu: 2	Obamba: 4	Civili: 1	Ikota: 3

74 of the students state that they have Gabonese languages as mother tongue. This is due to the fact that some Gabonese languages (Fang, Myene, Yipunu and Yinzebi) have the status of vehicular languages on a regional level, while other vehicular languages are domestic languages (cf. Idiata 2005). 26 of the students claim to have French as mother tongue, which results from the status of French in Gabon. In fact, the Gabonese languages co-exist with French, which is the only official language. French in Gabon is the language of knowledge and science, the language of power and rights, the language which ensures social and professional promotion (Mba-Nkoghe 1981: 23).

- Most Gabonese students had European languages as subjects. The distribution of the data is as follows:

French: 80	Italian: 5	Spanish: 39	Mandin: 1
English: 74	Portuguese: 3	German: 5	Latin: 2

From the results of the questionnaire, it may seem that Gabonese languages cannot be taken as subjects. However, this is not the case as the use of the Gabonese languages today is the result of the radical innovation in the educational language policy. In Gabon, particularly in Libreville, some secondary schools offer some Gabonese languages as subjects. This could be extended to primary schools, universities and to other educational institutions in Gabon.

- The Gabonese students were asked to list the languages they mainly speak at home. The distribution of the data is as follows:

Fang: 11	Lekanighi: 2	Yipunu: 2	Ikota: 1
French: 82	Yinzebi: 3	Lumbu: 1	Lembaama: 1
Tsogho: 1	Awandji: 1		

The majority of Gabonese students have studied French and mainly speak French at home. This is due to the privileged position and the official status of French in Gabon.

- The subjects studied by the Gabonese students at their respective universities are the following:

Lexicography: 6	Business Management: 22	Biotechnology: 5
Geology: 5	Conservation Ecology: 5	Sociology: 1
Forestry: 3	Phonetics: 1	Marketing: 3
English: 26	Mechanical Engineering: 4	Mathematics: 2
History: 2	Food Science: 1	Biochemistry: 1
Medical Science: 1	Architecture: 1	Economics: 1
Botany: 2	Zoology: 1	Physiology: 1
Food Science: 1	Environmental Health: 1	

- The question about the ability of the Gabonese students to write their mother tongues was answered as follows:

Yes: 48% No: 52%

As more than 50% of the Gabonese students cannot write their mother tongues, there is a need for learning to write the Gabonese languages. A dictionary serving as an example could be a solution to this. The function of dictionaries in text production should be taught to students. Compilers of future dictionaries in the Gabonese languages should be responsible for promoting the orthography of the Gabonese languages considered as standard or accepted by the majority of Gabonese language users in order to write efficiently in these languages.

4. Research methods

A questionnaire comprising 22 questions was given to all subjects. This questionnaire, based on the one developed by Hartmann (1999), was adapted for Gabonese users. Though different in objective, the method is comparable to that employed by Nesi and Hail (2002) who investigated the dictionary-using habits of international students studying through the medium of English at a British university.

5. Research coverage

The questionnaire given to participants in the survey dealt with the following 22 items:

- (1) First use and acquisition of a dictionary
- (2) Types of dictionaries owned

- (3) Number of dictionaries owned
- (4) Knowledge and use of electronic dictionaries
- (5) Types of dictionaries existing in the Gabonese languages
- (6) Types of dictionaries most frequently used in Gabon
- (7) Type of dictionary first to be compiled in the Gabonese languages
- (8) Type prioritized when buying a new dictionary
- (9) Reason for the last acquisition of a dictionary
- (10) Circumstances under which a dictionary is used
- (11) Reasons for using a dictionary
- (12) Frequency of using dictionaries
- (13) Use of information in the appendices of dictionaries
- (14) Awareness of the user guide notes in the front matter of dictionaries
- (15) Ability to find the information needed in a dictionary
- (16) Type of information the most difficult to find in dictionaries
- (17) Causes for difficulties in finding information in dictionaries
- (18) Satisfaction with the ability to use a dictionary
- (19) Opinions regarding the use of a dictionary
- (20) Attitudes towards the use of dictionaries
- (21) Instruction in dictionary use
- (22) Importance of being taught to use a dictionary

6. Research findings, description and analysis of the data

6.1 When did you start using and when did you acquire a dictionary?

61% and 58% of Gabonese students claimed to have started using and to have acquired a dictionary at primary school, 26% and 37% at secondary school and 0% and 2% at university.

This is an important question because it indicates when there should be started with teaching the use of dictionaries. Dictionaries are introduced at primary school in Gabon. Therefore, instruction in dictionary use should also start at this stage.

6.2 What types of dictionaries do you own?

The following were specified:

- General dictionary (e.g. Dictionary of English): 85%
- Special subject dictionary (e.g. Dictionary of Music): 18%
- Bilingual dictionary (e.g. English–French Dictionary): 81%
- Thesaurus (e.g. Dictionary of Synonyms): 36%
- Encyclopaedia: 46%

Gabonese students are very familiar with monolingual and bilingual dictionaries, but unfamiliar with special subject dictionaries. Dictionary typology explaining the main dictionary types should be taught at primary school.

6.3 How many dictionaries do you own?

19 students claimed to have 4 dictionaries
26 students claimed to have 3 dictionaries
27 students claimed to have 2 dictionaries
6 students claimed to have 1 dictionary

72 students have more than one dictionary, with only 6 saying that they own 1 dictionary. These results indicate that on the whole Gabonese students are dictionary knowledgeable.

6.4 Do you own any electronic dictionaries?

Yes: 21% No: 78%

As only a small number of Gabonese students own any electronic dictionaries, an introduction to electronic dictionaries and their use should be taught at school.

6.5 Which dictionary type(s) exist in the Gabonese languages?

The results indicating the answers to this question are as follows:

General dictionary (e.g. Dictionary of French): 36%
Special subject dictionary (e.g. Dictionary of Music): 10%
Bilingual dictionary (e.g. English–French Dictionary): 46%
Thesaurus (e.g. Dictionary of Synonyms): 38%
Encyclopaedia: 46%

As has already been stated above, the only type of dictionary existing in the Gabonese languages are bilingual dictionaries. Different answers given by students show that some do not know which specific dictionaries are available in the Gabonese languages. These results prove that Gabonese students have mostly been introduced to French dictionaries.

Existing dictionaries in the Gabonese languages are mainly unknown to the students. A course in the history of Gabonese lexicography, aiming to make known the available dictionaries in the Gabonese languages, should be taught at school.

6.6 Which dictionary type(s) do you use most frequently in Gabon?

The percentages obtained to this question are the following:

General dictionary (e.g. Dictionary of French): 79%
Special subject dictionary (e.g. Dictionary of Music): 9%

Bilingual dictionary (e.g. English–French Dictionary): 63%
Thesaurus (e.g. Dictionary of Synonyms): 36%
Encyclopaedia: 67%

This shows that most students prefer a general dictionary, followed by an encyclopaedia and a bilingual dictionary, and then a thesaurus, with only a small percentage selecting a special subject dictionary. Gabonese students are mostly using general dictionaries, encyclopaedias and bilingual dictionaries in French, the official language of Gabon. There is certainly a dictionary culture as far as French is concerned. Referring to Hausmann (1989: 13), Gouws and Prinsloo (2005: 12) stress that:

A dictionary culture implies that the members of a given speech community are familiar with different types of dictionaries and with the contents and presentation of these dictionaries. It also implies that they have reached a certain level of dictionary using skills. These skills do not come instinctively but need to be acquired.

The observation of Nyangone Assam and Mavoungou (2000: 226) is still valid: 'As far as the lexicographic needs of Gabonese languages are concerned, there is a lack of a dictionary culture.'

6.7 Which type of dictionary should first be compiled in the Gabonese languages?

The following results were obtained:

General dictionary (e.g. Dictionary of French): 32%
Special subject dictionary (e.g. Dictionary of Music): 2%
Bilingual dictionary (e.g. English–French Dictionary): 64%
Thesaurus (e.g. Dictionary of Synonyms): 17%
Encyclopaedia: 12%

Most students show a preference for a bilingual dictionary, then a general dictionary, followed by a thesaurus and encyclopaedia, and lastly a special subject dictionary.

The compilation of dictionaries in the Gabonese languages should start with bilingual ones. In Gabon where most of the languages are developing languages, bilingual dictionaries are seriously needed. As the Gabonese languages lack a solid lexicographic tradition, studying aspects of English and French lexicography as examples is inevitable. Problems of dictionary use in English and French could be regarded as a way of approaching problems likely to occur in the Gabonese languages. In this regard Mickala Manfoumbi (2004) took the courageous step of publishing the *Lexique pové–français / français–pové*, a bidirectional lexicon of one of the minority languages in Gabon. Future bilingual dictionaries of the Gabonese languages need to be planned carefully in order to meet the needs of Gabonese dictionary users.

6.8 What is your priority when buying a new dictionary?

- Its relevance to your needs: 90%
- The number of words: 12%
- The number of examples: 10%
- A reasonable price: 12%
- The reputation of the publisher: 16%
- Finding it convenient to be carried about: 3%

An overwhelming percentage of the Gabonese students indicated that they were led by their needs when deciding on which dictionary to buy. The standing of the publisher together with the number of words and examples and an affordable price received a much lower percentage. Gabonese students are aware that dictionaries are made to satisfy or respond to users' needs.

6.9 When you last bought a dictionary, was it ...?

- because a teacher or tutor recommended it: 23%
- because a friend or relative suggested it: 7%
- as a result of your own deliberate choice: 69%
- as a result of an advertisement: 0%
- due to an impulse: 8%
- I cannot remember: 3%

Most Gabonese students (69%) indicated that they made their own decision when buying a dictionary, while a much smaller number (23%) stated that they were influenced by the advice of a teacher or tutor.

To help students and teachers to come to a satisfactory decision when buying a dictionary, Gabonese lexicographers and booksellers should make useful information about dictionaries available, such as:

- *The age of the user:* It should be ascertained whether the dictionary is suitable for children, adolescents or adults.
- *The function of the dictionary:* It should be checked whether it is a monolingual or a bilingual dictionary, meant for text production or text reception, for mother-tongue speakers or learners, etc.
- *The language used in the dictionary:* The task of the compilers of dictionaries is to record the way the language is used. The dictionary should have included new words and expressions.
- *The number of lemmas (headwords), i.e. the number of words defined in the dictionary:* The user will nearly always be given this information on the cover.

- *The language of the definitions:* Users need to have a close look at the nature of the definitions and the words used in the definitions. Good dictionaries ensure that definitions are easy to understand, and that difficult words in the definition can also be found in the dictionary.
- *The amount of information given in each entry:* Grammatical information, pronunciation, spelling, etymology, example sentences, etc. should be included in the dictionary.
- *Illustrations:* Research has shown that illustrations do not necessarily help users (especially children) acquire vocabulary or understand meanings, although they do add to the visual appeal and 'brighten' the text.
- *Ease of reading and use:* The user must consider features such as the legibility of the typeface, the layout of the entries, and appendices such as spelling rules, common abbreviations, weight and measures, foreign phrases, and countries of the world, which can be useful additional information.

6.10 When do you use a dictionary?

- During class: 20%
- During an examination: 4%
- While studying at home: 91%
- While studying in a library: 59%
- Other (specify): 14%

The majority of Gabonese students (91%) make use of dictionaries when studying at home, and a lesser number (59%) when studying in a library. Only 20% use a dictionary during class and a small 40% use dictionaries during an examination. Allowance should also be made for students to use dictionaries during examinations.

6.11 Do you use a dictionary while you ...?

- read newspapers and magazines: 47%
- read textbooks: 65%
- read academic journals: 44%
- read a book for entertainment: 27%
- work on a translation exercise: 75%
- play word games: 45%

Almost three-quarters of the students use dictionaries while they work on a translation exercise or read textbooks. Nearly the same number use dictionaries when reading newspapers, magazines and academic journals. This is also the

case when they do crosswords and play word games, which shows that these activities encourage Gabonese students' use of dictionaries and improve their dictionary skills.

6.12 Do you use a dictionary ...?

- when you write: 63%
- when you read: 67%
- when you listen: 20%
- when you speak: 1%
- to look up a definition/equivalent of a word: 74%
- to look up the spelling of a word: 58%
- to look up synonyms/words of a similar meaning: 46%
- to look up examples of a word's use: 37%
- to look up a grammar point, e.g. part of speech: 20%
- to look up encyclopaedic information: 13%
- to look up the pronunciation: 41%

Most Gabonese students use a dictionary for finding the meaning of a word, followed by information on spelling, synonyms, pronunciation, usage, and grammar. Only a small number look for encyclopaedic information. Gabonese students use dictionaries for communicative purposes, i.e. encoding or decoding purposes: 63% when they write, 67% when they read. The importance of each data type should be emphasized in the teaching of dictionary use, particularly encyclopaedic information.

6.13 Do you ever use information contained in the appendices?

The students were exposed to a list of options including the most common information appendices found in different dictionaries. The percentages for the different appendices indicated were as follows:

- List of abbreviations: 57%
- List of regular/irregular verbs: 73%
- Proper names : 26%
- Units of measurement: 21%
- Other: 12%

From the above results, it becomes clear that students should be taught the structures of the dictionary, particularly the frame structure, consisting of the central list (the dictionary proper), the front matter (all functional text parts preceding the central list) and the back matter (texts coming after the central list) and how to retrieve the maximum information from this variety of texts.

6.14 If you are aware of the user guide notes in the front matter of the dictionary, do you ...?

study them: 24%
find them user-friendly: 31%
manage without them: 40%

These results show that 40% of the students ignore the information in the user guide notes, which is not very encouraging when it is considered that these provide useful information (such as the organization of the dictionary, an explanation of the pronunciation, etc.). Familiarity with the user guides could produce an immediate effect in both understanding and reducing the lookup process. The importance of the user guide notes should be taught to students.

6.15 Do you ever consult a dictionary without being able to find the information you need?

Very often: 4% Often: 11%
Sometimes: 76% Never: 10%

A large percentage of Gabonese students sometimes do not find the information they need. It is important to teach students how to obtain the required information.

6.16 What type of information is most difficult to find?

The percentages of the answers to this question are as follows:

General words: 2%
Specialized technical terms: 75%
Common words in a special subject area: 34%
Idioms, proverbs, phrases: 26%

Three-quarters of the Gabonese students selected specialized technical terms as the most difficult type of information to find. Just over one-third of the students and just under one-third of the students also indicated that they have difficulties with special subject terms and idioms, proverbs and phrases respectively. Gabonese students particularly need help on how to trace specialized technical terms in a dictionary.

6.17 What do you think are the causes of these difficulties?

My lack of dictionary skills: 13%
My lack of dictionary knowledge: 7%
Not enough information in the dictionary: 63%

The unclear layout of the dictionary: 22%
Because I do not read the instructions to the user: 13%

Gabonese students attributed the greatest part of their difficulties to the dictionary itself. In fact, the students believe that these problems are mainly the result of there not being enough information in the dictionary, while others related these problems to the unclear layout of the dictionary, a lack of dictionary skills, not reading the user instructions and a lack of dictionary knowledge were also seen as factors causing difficulties. From a user point of view, the importance of dictionary skills needs to be taught to the users to avoid difficulties in dictionary consultation.

6.18 Are you on the whole satisfied with your ability to use a dictionary?

Yes: 81% No: 19%

Four-fifths of the Gabonese students are satisfied with their ability to use a dictionary, while one-fifth feel that they have to acquire more dictionary skills.

6.19 In your opinion, using dictionaries is ...

easy: 48%	tedious/boring: 3%
difficult: 3%	worthwhile/informative: 58%
exciting/fun: 24%	of little help/not worth the trouble: 2%

Except finding the use of dictionaries worthwhile and informative (almost 60%), Gabonese students also consider the consultation process easy (almost 50%). This indicates that Gabonese students are familiar with dictionaries, also showing an awareness that dictionaries have a knowledge-orientated function.

6.20 Based on your experience, with which of the following statements do you agree?

Using dictionaries can improve my reading: 38%
Using dictionaries can improve my writing: 66%
Using dictionaries can help my speaking: 43%
Using dictionaries can help me perform better in my studies: 53%

About two-thirds of the Gabonese students state that using dictionaries can improve their writing, followed by about half who think that using dictionaries can help them perform better in their studies. Slightly less than half of the students think that using dictionaries can improve their reading and their speaking. Functions of dictionaries in text production and in text reception should be taught to students.

6.21 Have you ever been taught how to use a dictionary in Gabon?

Yes: 32% A little: 38% Never: 27%

As can be seen from these answers, almost 30% has never been instructed in dictionary use and almost 40% only slightly. These results show that teachers should insist more on the importance of the teaching of dictionary use in Gabon.

6.22 Do you think it is important for students in Gabon to be taught how to use dictionaries?

It is very important: 64% It is important: 25%
It is not important: 0% I do not know: 7%

Most Gabonese students (89%) think it is very important or important to be taught how to use dictionaries. As has already been stated, there is a need for user education in Gabon, especially for teacher training.

User education is defined by Hartmann and James (1998: 152) as 'the training of users in the reference skills in response to reference needs'. *Reference skills* are 'the abilities required on the part of the dictionary user to find the information being sought' (Hartmann and James 1998: 117), while *reference needs* are 'the circumstances that drive individuals to seek information in reference works such as dictionaries' (Hartmann and James 1998: 116). It should be agreed with Hadebe (2004: 90) that 'by training teachers to be good users themselves, it is envisaged that they could pass these skills on to students and eventually to society in general'.

With reference to the foregoing discussion, the choice of focusing on teacher training is justified. To improve the lexicographic situation in Gabon in general, Ekwa Ebanéga (2007) points out that the teaching of reference skills should first target the teachers themselves, for example those at the ENS (École Normale Supérieure) and the ENI (École Nationale des Instituteurs). Teachers should be taught reference skills as part of their training in order to enable them to make maximum use of dictionaries and reference books and impart this knowledge to students in high school. Teachers are in a better position than researchers or even lexicographers to assess the students' lexicographic needs. If teachers acquire dictionary skills and pass on these skills to students, a dictionary culture could develop in the community.

To introduce dictionaries and dictionary using skills (for example to teachers), Gouws (1996: 107) suggests that non-government organisations could be employed to give new and other literates access to dictionary culture. Dictionaries are compiled by lexicographers but produced and distributed by publishers and booksellers who should have the necessary expertise. Their assignment

is not only to make the dictionaries available but also to provide information on their optimal use. Publishing houses should co-operate with lexicographers and metalexicographers. Lexicographic planning for the Gabonese languages should take cognisance of these suggestions by Gouws.

Another approach to improve dictionary skills is the teaching of aspects of lexicography. Teachers should be initiated into the use of reference works of different types, dictionaries, thesauri and encyclopaedias, and others should receive basic training in dictionary structure, dictionary typology, lexicographic data (grammatical data, spelling, pronunciation, meaning, vocabulary etc.). There is also the need for teaching according to the level (beginner–intermediate–advanced) of their students.

In addition to the foregoing, teachers should be introduced into the history of the lexicography of the African languages in general and the Gabonese languages in particular. Missionaries and colonial administrators were the first who compiled dictionaries in the Gabonese languages. Similar examples have been observed in other African countries (cf. Mpofu 2001: 244, Mavoungou 2001: 123 and Hadebe 2004). Furthermore, teachers have to explain to users the motivations of the existing dictionaries. For example, with regard to external motivation, teachers could state that existing dictionaries in the Gabonese languages were not directed at specific speakers of these languages. As far as internal motivation is concerned, teachers could explain to users whether the dictionary is directed at promoting language or focusing on linguistic issues. For example, with regard to existing dictionaries in the Gabonese languages, teachers could explain to users the aim of these dictionaries. Knowing the history of the lexicography of the Gabonese languages (types of existing dictionaries, their different motivations, etc.) will help users to appreciate their own reference works.

Conclusion

The research discussed in this article has presented the viewpoints on dictionaries of a number of Gabonese students. Gabonese students start using and acquiring dictionaries in primary school. They most frequently use printed monolingual and bilingual dictionaries. They are less familiar with special subject dictionaries and electronic dictionaries. When they buy a new dictionary their priority is its relevance to their needs. They use a dictionary when they study at home and while they do translation exercises and read textbooks. They do not consult a dictionary in search of encyclopaedic or etymological information, but for the definition or equivalent of a word, the part of speech, pronunciation, spelling, synonyms, idioms or proverbs, and example sentences. For them, special technical terms are the most difficult information types to find in a dictionary. In their opinion, using dictionaries are easy and informative, can improve their reading, speaking and writing, and help them perform better in their studies. There is a dictionary culture as far as French is con-

cerned, but there is a lack of a dictionary culture with regard to the Gabonese languages. Gabonese language dictionaries compiled by missionaries and colonial administrators should be known by and accessible to users. Bilingual dictionaries for text production and/or text reception aimed at Gabonese users should be the first to be compiled in order to establish a dictionary culture in the Gabonese languages.

The use of Gabonese language dictionaries presupposes that the student knows the alphabetical ordering principle of the dictionary and is acquainted with the use of keywords at the top of the pages. It also presupposes a basic knowledge of the grammatical structure of the language. Mother-tongue education is needed and required for the use of Gabonese language dictionaries.

The importance of instruction in dictionary use is not underestimated by Gabonese students. It is important for them to be taught how to use a dictionary, how to find the required information, and how to retrieve the maximal information in the variety of texts contained in a dictionary (front matter, central list, back matter). They want to be taught the functions of dictionaries (text production or text reception), to be shown the importance of the user guide notes, to be made aware of the variety of dictionaries available and the differences between them, and to be guided to choose the most suitable one when buying a dictionary. Similarly, Gabonese teachers should be taught how to use a dictionary. As far as Gabonese users are concerned, instruction in dictionary use should be given at primary level and continued at secondary and university levels.

This survey is a preliminary study as far as the Gabonese user profile is concerned. It is not exhaustive. Further quantitative and qualitative research and similar experiments need to be conducted in Gabon with reference to different dictionary users (teachers, children, adolescents and different speech communities), dictionary types, content, presentation of dictionaries, dictionary use, etc.

Bibliography

- Atkins, B.T.S. and K. Varantola. 1998. Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. Atkins, B.T.S. (Ed.). *Using Dictionaries*: 21-81. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- 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.
- Bogaards, P. 1988. À propos de l'usage du dictionnaire de langue étrangère. *Cahiers de lexicologie* 52(1): 131-52.
- De Schryver, G.-M. and D.J. Prinsloo. 2000. The Compilation of Electronic Corpora, with Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.
- Ekwa Ebanéga, G.-M. 2007. *A Microstructural Programme for Dictionaries in Fang*. Unpublished D. Litt. Dissertation. Stellenbosch: University of Stellenbosch.

- Gouws, R. H.** 1989. *Leksikografie*. Pretoria: Academica.
- Gouws, R.H.** 1996. A Sequence for Meeting Lexicographic Needs. *Lexicography as a Financial Asset in a Multilingual South Africa*: 97-110. Pretoria: Department of Arts, Culture, Science and Technology.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hadebe, S.** 2004. Improving Dictionary Skills in Ndebele. *Lexikos* 14: 89-104.
- 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.** 1999. Case Study: The Exeter University Survey of Dictionary Use. Hartmann, R.R.K. (Ed.). 1999. *Dictionaries and Language Learning: Recommendations, National Reports and Thematic Reports from the TNT Sub-project 9*. <http://web.fu-berlin.de/elc/tnp1/SP9dossier.doc>.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hausmann, F.J.** 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, F.J. et al. (Eds.). 1989-1991: 1-19.
- 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*. Berlin: De Gruyter.
- Householder, F.W. and S. Saporta (Eds.).** 1962. *Problems in Lexicography*. Bloomington: Indiana University/The Hague: Mouton.
- Idiata, F.** 2005. *Les langues du Gabon: données en vue d'une classification fondée sur le critère d'intercompréhension*. Cape Town: Centre for Advanced Studies of African Society (CASAS).
- Mavoungou, P.A.** 2001. Macro- and Microstructural Issues in Mazuna Lexicography. *Lexikos* 11: 122-138.
- Mba-Nkoghe, J.** 1981. Le Gabon linguistique. *Notre librairie. Littérature gabonaise* 105: 20-23.
- Mickala Manfoumbi, R.** 2004. *Lexique pove-français/français-pove*. Libreville: Éditions Raponda-Walker.
- Mpofu, N.** 2001. Problems of Equivalence in Shona-English Bilingual Dictionaries. *Lexikos* 11: 242-251.
- Nesi, H. and R. Haill.** 2002. A Study of Dictionary Use by International Students at British University. *International Journal of Lexicography* 15(4): 277-305.
- Nyangone Assam, B. and P.A. Mavoungou.** 2000. Lexicography in Gabon: A Survey. *Lexikos* 10: 252-274.
- Tarp, S.** 2002. Translation Dictionaries and Bilingual Dictionaries: Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Tomaszczyk, J.** 1979. Dictionaries: Users and Use. *Glottodidactica* 12: 103-19.
- Whitcut, J.** 1986. The Training of Dictionary Users. Ison, R. (Ed.). *Lexicography. An Emerging International Profession*: 111-122. Manchester: Manchester University Press.
- Wiegand, H.E.** 1998. Altes und Neues zur Makrostruktur in alphabetischen Printwörterbüchern. Wiegand, H.E. (Ed.). 1998. *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*: 348-372. *Lexicographica. Series Maior* 84. Tübingen: Max Niemeyer.

L.T. Marole: A Forgotten Pioneer in Tshivenda Lexicography*

Munzhedzi James Mafela, *Department of African Languages,
University of South Africa (UNISA), Pretoria, Republic of South Africa
(mafelmj@unisa.ac.za)*

Abstract: L.T. Marole is a pioneer in Tshivenda lexicography. He is especially known for the compilation of word lists and phrase books. Through his production of these lexicographic works, Marole has preserved some Tshivenda terminology which is now regarded as obsolete; for example, terminology regarding names of animals, parts of the body, kinship relations and sicknesses. This type of terminology could be useful to the present younger generation. Readers may learn much about Tshivenda culture from Marole's works. However, his works have not received recognition from Vhavana dictionary users and compilers. The public is largely unaware of his works. This article aims at making the public aware of Marole's contribution to the development of Tshivenda lexicography, emphasising that this contribution should not be ignored as being insignificant.

Keywords: LEXICOGRAPHY, TRANSLATING DICTIONARY, BILINGUAL DICTIONARY, LINGUISTICS, TERMINOLOGY, LEXICAL ENTRY, EQUIVALENT, TRANSLATOR, CULTURE, KINSHIP TERMS, ORTHOGRAPHY, CONTEXT, DATABASE, DIALECT

Opsomming: L.T. Marole: 'n Vergete pionier in Tshivendaleksikografie.

L.T. Marole is 'n baanbreker op die gebied van Tshivendaleksikografie. Hy is veral bekend vir die samestelling van woordelyste en taalgidse. Deur sy produksie van hierdie leksikografiese werke, het Marole sekere Tshivendaterme wat nou as verouderd beskou word, bewaar; byvoorbeeld, terminologie betreffende die name van diere, liggaamsdele, verwantskapsbetrekkinge en siektes. Hierdie soort terminologie sou nuttig kon wees vir die huidige jonger geslag. Lesers kan baie omtrent die Tshivendakultuur uit Marole se werke leer. Sy werke het egter nie erkenning gekry van Vhavenda woordeboekgebruikers en -samestellers nie. Die publiek is grootliks onbewus van sy werke. Hierdie artikel beoog om die publiek bewus te maak van Marole se bydrae tot die ontwikkeling van Tshivendaleksikografie deur te beklemtoon dat hierdie bydrae nie as onbelangrik geïgnoreer behoort te word nie.

Slutelwoorde: LEKSIKOGRAFIE, VERTALENDE WOORDEBOEK, TWEETALIGE WOORDEBOEK, LINGUISTIEK, TERMINOLOGIE, LEKSIKALE INSKRYWING, EKWIVALENT, VERTALER, KULTUUR, VERWANTSKAPSTERME, ORTOGRAFIE, KONTEKS, DATABASIS, DIALEK

* This article was presented as a paper at the Eleventh International Conference of the African Association for Lexicography, organized by the Tshivenda National Lexicographic Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5–7 July 2006.

Introduction

Since spoken Tshivenda was reduced to writing by the Berlin missionaries in the late nineteenth century, little has been achieved in the field of Tshivenda lexicography. Mawela (1999: 151) writes: 'At present there are only a few dictionaries, all of them translating dictionaries ...' The basic aim of a translating dictionary is to coordinate lexical units of one language with those units of another language which are equivalent in their lexical meaning (Zgusta 1971). At the time of the writing of these word lists and phrase books, the main need of missionaries, students and other foreign speakers was to learn Tshivenda. This is emphasised by Wentzel and Muloiwa (1982) in their preface comment on the purpose of their *Improved Trilingual Dictionary*:

The original edition of the Trilingual Dictionary was the result of the need to provide those students who follow the Special Course in Venda with a practical list. After publication it was found, however, that the interest in and need for such a work was much wider than anticipated and that the basic need pointed towards a small practical dictionary.

The special course referred to is directed mainly at speakers of Tshivenda as a second language. Translational dictionaries are important in the development of languages, especially the indigenous languages of the Republic of South Africa, as these languages come into contact with different cultures. As a result of the increase in international interaction, more and more people are beginning to realize that bilingual dictionaries have become necessary economically, intellectually and culturally (Al-Kasimi 1983).

Mathivha (1972: 104) has this to say about the development of Tshivenda lexicography:

The development of dictionaries in the Venda language was started by Rev. E. Gottschling between 1897 and 1903 by making a collection of words written on ruled pages. This dictionary, however, has never been published.

The first linguists who made a contribution to Tshivenda lexicography were Th. and P. Schwelnus who published Tshivenda-German word lists in 1904 and 1918, the first dealing with Tshivenda verbs only. However, it was not before 1932 that L.T. Marole's first English-Tshivenda phrase book was published, followed in 1936 by an English-Tshivenda word list, co-compiled with F.J. de Gama. Shortly afterwards, in 1937, N.J. van Warmelo's Tshivenda-English dictionary appeared. In 1954, the third edition of Marole and De Gama's English-Tshivenda vocabulary was published. This was followed in 1955 by Marole's two phrase books in English-Tshivenda and Afrikaans-Tshivenda. In 1958, under the editorship of Van Warmelo, a Tshivenda terminology for school use appeared. A Tshivenda terminology and spelling list was published in 1962 by the then Department of Bantu Education. The years 1976 and 1982 saw the publication of the first and revised editions of P.J. Wentzel and T.W.

Muloiwa's trilingual dictionary: Tshivenda–Afrikaans–English. A revised Tshivenda terminology and orthography by the later Department of Education and Training appeared in 1980. Other Tshivenda dictionaries published during the nineteen-eighties were those by M.C. Neluvhalani (1987) and N.J. van Warmelo (1989). One of K.B. Harthorne's dictionaries of basic English which appeared in 1984 contains an English–Tshivenda word list. Only in 2006, the Tshivenda National Lexicography Unit published its *Thalusamaipfi*. Recently the Department of Arts and Culture has also produced a few terminology lists for specific subjects. Although some of these lexicographers were linguists, many of them had not been trained lexicographers. They became involved in lexicography out of interest. As can be seen from this summary of Tshivenda lexicographic publications, L.T. Marole was a pioneering lexicographer of Tshivenda. The aim of this article is therefore to highlight Marole's contribution to the development of Tshivenda lexicography as it has not received the recognition it deserves.

L.T. Marole's contribution to Tshivenda lexicography

L.T. Marole was the first Muvenda to publish a lexicographic work in Tshivenda. Up to 1955, four lexicographic works had been produced by Marole, i.e. *Phrase Book for English and Venda* (1932), *English–Tshivenda Vocabulary* (1936) which he co-compiled with F.J. de Gama and of which the third edition *English–Venda Vocabulary* appeared in 1954, *Phindulano: English–Venda Phrase Book* (1955) and *Afrikaans–Venda Vocabulary and Phrase Book* (1955). Though the date of publication of *Phindulano* does not appear in the book, according to Mathivha (1972), it was published in 1955 because it is claimed that it was published a year after *English–Venda Vocabulary*. Prof. D. Ziervogel received it as a complimentary copy in 1956. These lexicographic books are small in size but very valuable to foreign speakers who want to learn Tshivenda, and also to Vhavenda who intend to learn English and Afrikaans. They serve as a database of Tshivenda terminology. Marole's work is important because it preserves Tshivenda terminology not regularly used by the Vhavenda today. But although Marole has preserved the Tshivenda vocabulary and pioneered Tshivenda lexicography, he is hardly mentioned by Vhavenda lexicographers in their discussions of lexicographic matters. Dictionary users hardly use his works in their daily reference to dictionaries because they are unknown and unavailable. Even students of lexicography rarely mention his works in their research. Efforts should be made to make these lexicographic works available and familiar to the public again.

Marole was encouraged by the Berlin missionaries to write the *Phrase Book for English and Venda* which was meant for foreign language speakers, especially missionaries who wanted to learn Tshivenda at that time. Entries in the translational dictionary are arranged according to themes, i.e. parts of the body, wild animals, domestic animals, prepositions, seasons of the year, relation-

ships, sickness, housework, and interrogative pronouns. Grouping lexical entries according to themes is an effective way of teaching foreigners the language as learners are able to concentrate on one aspect of the language at a time.

At present, this type of dictionary could be useful to Vhavenda youth, especially those who live in urban areas and know little about Tshivenda names of animals, both wild and domestic; kinship relations in Tshivenda; Tshivenda names for various diseases, names of different types of housework and Tshivenda names for different parts of the body. Mathivha (1972: 105) says that the inclusion of names of plants and insects is a very important contribution to Tshivenda since some of these plants and animals are gradually disappearing. In the following paragraphs, a few examples illustrating the importance of Marole's lexicographic works in the preservation of Tshivenda terminology are given and discussed.

Few young Vhavenda know what the Tshivenda equivalent for **mouth** is. If one asks a Muvenda youth to provide the Tshivenda word for **mouth**, one may receive the answer *mulomo*. According to Marole (1932: 3), the Tshivenda equivalent of **mouth** is *hanwani*, and **lips** are *milomo*. One can speak of *mulomo wa n̄tha*, referring to the upper lip, and *mulomo wa fhasi*, referring to the lower lip. Again, modern young Vhavenda refer to both leg and foot as *mulenzhe*. Marole differentiates between these two parts of the body as *mulenzhe* and *lunayo* respectively. The names of animals such as **leopard**, **tiger** and **tiger-cat/serval** are confusing to many Vhavenda. However, Marole differentiates between them. In Tshivenda, according to Marole (1932), a leopard is *nngwe*, a tiger is *lugwen̄tofu*, and a tigercat/serval is *didingwe*. Young Vhavenda are not confused by the English terminology, being well aware of the corresponding English names. Few Vhavenda know what **breakfast** is in Tshivenda. A possible answer they might give is a phrase such as *zwīliwa zwa matsheloni* (morning food). Marole provides *vhuragane* as the Tshivenda equivalent of **breakfast**. When terms such as **queen** are to be translated, it is difficult to find an equivalent in Tshivenda. Sometimes translators may go to the extent of translating by explanation, whereas Marole provides an equivalent to **queen**, *Vhomakhadzi*. This is not an ordinary *makhadzi* (aunt). The honorific prefix *Vho-* indicates that it is a *makhadzi* possessing power.

In addition to enabling users of these lists to find the Tshivenda terms for the above-mentioned concepts, they will also learn about the culture of the Vhavenda. In Tshivenda culture, **brother** is denoted by two words, i.e. *murathu* and *mukomana*. Marole is able to distinguish between these by indicating that **younger brother** is *murathu*, and **elder brother** *mukomana*. If one asks a child how he/she is related to his/her brother, one may be surprised to learn that the child will answer **uncle**. Vhavenda children do not fully understand what the concepts **aunt**, **uncle**, **cousin**, etc. mean in Tshivenda. They attach English meanings to these terms, which in many instances are different from those in Tshivenda. Marole has successfully dealt with kinship relations in Tshivenda in

his lexicographic works which could be of great assistance in solving relational problems.

Some of the entries are illustrated by phrases at the end of a particular theme. Learners are shown how to use the entry words in sentences. For example (Marole 1932: 12):

My father is seriously ill. *Khotsi yanga ulwala nga maanḁa.*
What with? *Uvhulawa ngani?*
With coughing and headache. *Ndi mphigela na fhoho.*

In the list of the entries covering sicknesses, the equivalents of **coughing** and **headache** are provided as *uhogola* and *fhoho i a rema* respectively. In some cases, word-divisions are incorrect. For example, the sentence *Khotsi yanga ulwala nga maanḁa* should read *Khotsi anga u lwala nga maanḁa*. This is supported by Mathivha (1972: 106) who says: 'The Venda used in this book does not conform to the accepted Venda orthography and word-division.' However, a learner learns how to use the entries in sentences. One of the most important characteristics which helps dictionary users to understand lexical items is context. Dictionaries provide example sentences or phrases to illustrate meaning, grammar or usage. According to Cowie (1989: 57), illustrative examples may serve one or more of the following purposes: to supplement information in a definition; to show the entry word in context; to distinguish one meaning from another; to illustrate grammatical patterns; to show other typical collocations; and to indicate appropriate registers or stylistic levels.

Phindulano: English–Venda Phrase Book consists of two parts, i.e. a translation of phrases and a bilingual glossary. The translation section includes themes such as salutation, walking and riding, correspondence, weather conditions, buying and selling, hunger and thirst, getting up in the morning, coming home to sleep, about Vhaventḁa, going to the chief's place, idioms, teaching, etc. From these, learners not only acquire the meaning of different phrases in Tshiventḁa, but they also learn about the life of the Vhaventḁa. For example (Marole 1955: 16):

In former times, how did the Bawenda produce light, when they had no candles nor lamps? Nḁa kale kaleni Vhaventḁa vhovha vha tshi vho netshela nga mini unga makhandela na malamba vhovha vhasina?

They used dry tambookies grass for light. Vha shumisa hatsi ha madadasenga uita luvhone.

A dictionary user learns how the Vhaventḁa of the past used to make light in their houses. As in the *Phrase Book for English and Venda*, entries in the bilingual glossary are arranged according to themes, e.g. wild animals, birds and domestic animals. *Phindulano: English–Venda Phrase Book* could be of great help to Vhaventḁa youth who wish to acquire the correct vocabulary regarding the names of wild animals, birds and domestic animals.

The *English–Venda Vocabulary* is Marole's greatest contribution to the development of Tshivenda lexicography. This dictionary was co-compiled with F.J. de Gama. Unlike Marole's other lexicographic works, which are translational in nature, this dictionary provides parts of speech in its main entries. The following are some of the entries found in the dictionary (Marole 1954: 4):

- Adversary** (n.) swina, muhanedzi
Adversity (n.) tsiku, vhushayi, vhuṭungu
Advice (n.) ngeletshedzo
Advise (v.) u eletshedza
Advocate (n.) muimeleli, muambeli

The letters *n.* and *v.* indicate noun and verb respectively. Dictionary users are therefore in a position to learn the parts of speech of the lexical entries. In some instances, multiple equivalents have been provided for one entry as illustrated by the following examples (Marole 1954: 13, 15 and 18 respectively):

- Dear** (adj.) tshiduraho, (n.) mufunwa
Delegate (v.) uruma, (n.) muḍinda
Elect (v.) ukhetha, (n.) mukhethwa

The provision of multiple equivalents is of great assistance to dictionary users in discriminating between various meanings of words.

Tshivenda consists of more than one dialect, with Tshiphani as the standard dialect. Almost all dialects are catered for in Marole and De Gama's dictionary. A dictionary should not concentrate on standard language only; all dialects of the language should be included. For example (Marole 1954: 63):

- Sugar-cane** (n.) mphwe, mwovha

Sugar-cane is *mphwe* in Tshiphani, whereas in the dialect Tshiilafuri it is known as *mwovha*. Unlike Marole's other lexicographic works, these entries are not arranged according to themes, but alphabetically, following the style employed in modern lexicography.

Marole not only produced lexicographic works for English speakers who wanted to learn Tshivenda: Afrikaans-speaking users have also been catered for. In 1955, he produced *Afrikaans–Venda: Vocabulary and Phrase Book*. This contribution consists of two parts, i.e. a translation of phrases and a bilingual glossary. The entries have been arranged alphabetically. Days of the week, months of the year, and numbers are also included. Regarding numbers, he has used the traditional method of counting which is presently used by uneducated people living in rural villages. The following is an example of numbers as provided by Marole (1955: 11):

- 6 **ses** Nṭhanu na nthihi
7 **sewe** Nṭhanu na mbili

8	ag	Nṭhanu na ntharu
9	nege	Nṭhanu na ṛa
11	elf	Fumi na nthihi

According to educated people, *ses*, *sewe*, *ag*, *nege* and *elf* are known in Tshivenda as *rathi*, *sumbe*, *malo*, *ṭahe* and *fumithihi* respectively. The modern method of counting, which seems to be economical regarding space and time, is not used by ordinary Vhavaṅḁ because it is difficult to understand. For example, *rathi* (*ses*) is a coined numeral derived from the verb stem *-ratha*, similar to Sesotho sa Leboa *tshela*, both of which means *step over*. When one counts using the fingers, one will count the number *six* by stepping over to the other hand. However, the traditional way of counting is not economical, but simpler to understand. For example, counting the number *ses* (*six*), one would say *five and one* (*nṭhanu na nthihi*) as shown in the examples given above. Vhavaṅḁ youth should be exposed to this type of counting rather than using the coined expressions.

Conclusion

Marole's lexicographic work is not perfect. There are many errors in spelling and word-division. However, Marole's lexicographic products should not simply be forgotten as they are valuable both for modern lexicographers as well as learners of Tshivenda. Lexicographers should use works such as these as sources when compiling their dictionaries. Learners, especially the youth, will be exposed to quality vocabulary which is presently being neglected in favour of borrowings from other languages such as English.

N.J. van Warmelo's dictionary, *Tshivenda-English Dictionary*, published in 1937, was revised with the help of W.M.D. Phophi and published as *Venda Dictionary: Tshivenda-English* in 1989. Wentzel and Muloiwa's dictionary, *Trilingual Elementary Dictionary: Venda-Afrikaans-English*, published in 1972, was revised by the authors and published as *Improved Trilingual Dictionary: Venda-Afrikaans-English* in 1982. It may be possible to revise Marole's lexicographic works in a similar manner in order to preserve them and encourage their use by the greater public.

Marole's works could further form a base from which different types of dictionaries could be produced, for example, a dictionary of parts of the body, a dictionary of animals (both domestic and wild), a dictionary of idioms, a dictionary for tourists, etc. Dictionaries of this nature could prove useful for the youth as well.

It is therefore strongly recommended that an initiative is taken in revising Marole's lexicographic works. Marole never received any financial support from the government or any other private organization, yet he produced excellent work which is an asset to the Tshivenda language.

References

- Al-Kasimi, A.M.** 1983. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Cowie, A.P.** 1989. The Language of Examples in English Learners' Dictionaries. James, G. (Ed.). *Lexicographers and their Works*. Exeter: University of Exeter.
- Department of Bantu Education.** 1962. *Venda. Terminology and Spelling No. 2*. Pretoria: The Government Printer.
- Department of Education and Training.** 1980. *Venda. Terminology and Orthography No. 3*. Pretoria: The Government Printer.
- Hartshorne, K.B.** 1984. *Dictionary of Basic English: Venda*. Assisted by J.H.A. Swart and E. Posselt. Johannesburg: Educum.
- 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.** 1955. *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.
- Mathivha, M.E.R.** 1972. *A Survey of Literary Achievements in Venda: A Descriptive Study from the Earliest Beginnings up to 1970*. Unpublished D.Litt. et Phil. Thesis. Pietersburg: University of the North.
- Mawela, A.** 1999. The State of Tshivenda Lexicography. *Lexikos* 9: 251-254.
- Neluvhalani, M.C.** 1987. *Ifa Lashu Ila Maambole*. Johannesburg: Sasavona.
- Schweltnus, Th. and P.** 1904. Die Verba des Tšivenda. *Mitteilungen des Seminars für Orientalische Sprachen an der Königlich Friedrich-Wilhelms-Universität zu Berlin* 7(3): 12-31.
- Schweltnus, Th. and P.** 1918. Wörterverzeichnis der Venda-Sprache. *Jahrbuch der hamburgischen wissenschaftlichen Anstalten. Beiheft* 36: 51-78.
- Tshikota, S. (Ed.-in-chief).** 2006. *Tshivenda–English thalusamaipfi ya u thlutshedza ya nyambombili/English–Tshivenda Bilingual and Explanatory Dictionary*. Cape Town: Phumelela Publishers.
- Van Warmelo, N.J.** 1937. *Tshivenda–English Dictionary*. Department of Native Affairs. Ethnological Publications Vol. VI. Pretoria: Government Printer.
- Van Warmelo, N.J. (Ed.).** 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: Luvenḁa–Luwuhuru–Luisimane/Verbeterde drietalige woordeboek: Venda–Afrikaans–Engels/Improved Trilingual Dictionary: Venda–Afrikaans–English*. Pretoria: University of South Africa.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

Bibliographical Materials for Afrikaans Etymological Lexicography

Jeremy Bergerson, *Department of German, University of California,
Berkeley, United States of America (jbergerson@berkeley.edu)*

Abstract: Afrikaans etymological lexicography has yet to see an analytic dictionary along the lines of Sigmund Feist's *Gotisches etymologisches Wörterbuch* or Anatoly Liberman's *An Analytic Dictionary of English Etymology*. Before an analytic dictionary of Afrikaans etymology can be written, a comprehensive bibliography of works and articles on the subject must first be brought together. This collection is a first step towards creating a resource for Afrikaans similar to Liberman's *A Bibliography of English Etymology*.

Keywords: ETYMOLOGY, LEXICOGRAPHY, HISTORICAL LEXICOLOGY, DICTIONARIES, WORD STUDIES, BIBLIOGRAPHY, ANALYTIC, AFRIKAANS, HISTORICAL LINGUISTICS

Opsomming: Bibliografiese materiaal vir die Afrikaanse etimologiese leksikografie. Die Afrikaanse etimologiese leksikografie moet nog 'n analitiese woordeboek kry min of meer soos Sigmund Feist se *Götisches etymologisches Wörterbuch* of Anatoly Liberman se *An Analytic Dictionary of English Etymology*. Voordat 'n analitiese woordeboek van die Afrikaanse etimologie geskryf kan word, moet 'n omvattende bibliografie van werke en artikels oor die onderwerp eers bymekaargebring word. Hierdie versameling is 'n eerste stap tot die totstandbrenging van 'n hulpmiddel vir Afrikaans soortgelyk aan Liberman se *A Bibliography of English Etymology*.

Sleutelwoorde: ETIMOLOGIE, LEKSIKOGRAFIE, HISTORIESE LEKSIKOLOGIE, WOORDEBOEKE, WOORDSTUDIE, BIBLIOGRAFIE, ANALITIES, AFRIKAANS, HISTORIESE TAALKUNDE

As etymologies usually appear in dictionaries, one would expect below a discussion of the history of Afrikaans lexicography. The publication of Gouws and Ponelis,¹ however, has rendered another detailed treatment of the subject unnecessary. Nevertheless, bearing in mind the somewhat obscure development of Afrikaans, the earliest commentaries upon Afrikaans-Hollands, or Cape Dutch,² assume importance as starting points in the history of Afrikaans etymology. Given the scope of this project, moreover, it seems worthwhile to provide some background to the materials handled here.

In the course of the eighteenth century and into the nineteenth, travellers, botanists, and foreigners in the Dutch East India Company's employ remarked upon the local speech at the Cape of Good Hope. Since none of these sojourners

were trained linguists, they tended not to comment on phonological or otherwise arcane linguistic matters. Instead, they focused on the most noticeable of deviations, word usage.³ Their numerous tracts and travelogues are of great value to the etymologist, providing as they do important material for the dating of individual words. J. du P. Scholtz⁴ excerpted all of these texts to write 'Leksikologiese aantekeninge uit ou Kaapse bronne'. It remains the most authoritative study to date.

Starting in the mid 1820s and continuing through the rest of the century one comes across numerous letters and articles written in varieties of Afrikaans and appearing in various newspapers (among others *De Zuid-Afrikaan* and *Het Kaapsche Grensblad*). These writings did not go unnoticed, and in Cape publications such as *The South African Magazine* and *The Cape Monthly Magazine*, polemical as well as expository articles on 'Cape Dutch' followed that drew heavily on particular words to shore up vying explanations. The participants in these debates were, among others, Pres. F.W. Reitz (then Chief Justice of the Republic of the Free State) and M.L. Wessels (a Cape advocate), and despite their learned tenor and sometimes accurate speculations, these disputations, though they proved influential in spurring on later scholars, are on the whole unscientific.⁵

The earnest study of the Afrikaans lexicon begins with A.N.E. Changuion's 'Proeve van Kaapsch Taaleigen' which appeared as a supplement to the second edition of his *Nederduitsche taal in Zuid-Afrika hersteld*. It was, by the author's own admission, intended to provide some linguistic value,⁶ but the work is dominated by pedantic caveats.⁷ This and seven other early and particularly useful word lists (spanning the years 1848–1908) were brought together in Van der Merwe (1971). Among these early attempts at describing the Afrikaans vocabulary is *Proeve van een Kaapsch-Hollandsch idioticon* by N. Mansvelt. His little 'idioticon' contains a goodly amount of information from previous word lists, and provides interesting notes and etymologies of its own. Its effect upon the thinking of later scholars (Hugo Schuchardt, D.C. Hesseling) together with its numerous insights makes it one of the principal texts in Afrikaans etymology. But in spite of all the interesting conjecture and analysis represented in this body of work, none of it ever reached the standard by D.C. Hesseling.⁸

The polemic between Hesseling and D.B. Bosman, and tangentially P.J. du Toit, is the first serious debate about the origin of Afrikaans and forms the basis of scholarly literature on the subject. This had great influence on S.P.E. Boshoff, whose dissertation *Volk en taal van Suid-Afrika* is indispensable to the etymologist. Boshoff went on to author *Etimologiese woordeboek van Afrikaans*, and later to co-author *Afrikaanse etimologieë* with G.S. Nienaber. For the rest of the lexicographical-etymological history of Afrikaans, see Gouws and Ponelis.

Deciding which etymologies to include in this bibliography proved difficult. For example, one often comes across a statement like '*baie* (< Mal. *banjak*)'. Here the author has identified the parent language of the word *baie*, but has failed to provide a discussion. This is not an etymology. Such instances are

simply dealt with, however, as was the case in two particular works on Afrikaans words,⁹ some etymologies seemed promising, but proved too questionable to enter here. In moments of unsurity, common sense prevailed; e.g. if the author was a main contributor to the field (Abel Coetzee, G.S. Nienaber, J. du P. Scholtz), he was rather included in the bibliography than not. So the number of useless (uninformed, trivial, undeveloped) etymologies has been kept to a minimum.

Deciding which words to enter was easier: the principles behind choosing which words to include in the present bibliography are consistent with those employed in *Afrikaanse etimologieë*. That is, words shared with Dutch, such as *broer*, *praat*, or *stoel* are left out because they have already been dealt with by Dutch etymologists. However, some words that are clearly Dutch but have a meaning unique to Afrikaans have been included, like *die* 'the' (Dutch 'those'; relative pronoun), *lemoen* 'an orange' (Dutch 'lemon') and *seun* 'boy' (Dutch 'son'). Also, all common Afrikaans words that are dialectal in Dutch are incorporated into the list, e.g. *dalk* 'maybe,' *kiewiet* 'a type of bird,' and *stadig* 'slow.' Some words have enjoyed much attention because they have a phonetic form that is significantly different from Dutch: *die* 'the,' *het* 'have,' *hom* 'him,' *hulle* 'they, them, their,' *watter* 'which.' These have also been included. Some words are well-represented (*baie*, *ons*), and many others have only one reference. This is due to the fact that the majority of the best etymologies are to be found in books, see the list below. The entry of foreign words has been kept to a minimum, unless they were either words that showed up repeatedly in the literature or words that comprise an integral part of the Afrikaans vocabulary.¹⁰ The list contains a few obsolete words, such as *jaarhonderd* 'century,' *sulle* 'they,' and *tof* 'smart, dignified.' As a rule, words have been included only when the place of their origin is ascertainable. Since Scholtz supplied important comments regarding the origin or nature of words in Afrikaans, his works have been excerpted even when he did not provide etymologies. Plant and animal names were generally not added, being so numerous as to merit a separate study. Though the scope of this bibliography has been limited to works in Afrikaans, a few important Dutch, English, and German articles have been added, without which the bibliography would have been incomplete. But for a few exceptions, newspaper articles have been omitted.

The definitions in the word list are meant for identification only. Here I followed the format of the *Bibliography of English Etymology* being prepared at the University of Minnesota. To determine the spelling of the words here, the *Afrikaanse woordelys en spelreëls* and other authoritative dictionaries such as the *Woordeboek van die Afrikaanse Taal* and *Verklarende Handwoordeboek van die Afrikaanse Taal* were consulted.

A few articles about the Afrikaans vocabulary do not contain etymologies but are useful to the etymologist:

Du Plessis-Müller, S.F. 1938–39. Die woordeskat van die Dagboek van Louis Trichardt. *TWK* 17: 31-9, 56-9

- Du Toit, S.J. 1938. Die Gordon-versameling. *TWK* 16: 108-11
- Muir, John. 1929. Gewone plantname in Riversdal. *Huisgenoot* 30 Augustus: 45, 47, 49, 51, 55
- Paardekooper, P.C. 1990. Hollandse zeemanstaal(?) en Afrikaanse waltaal. *TNTL* 106: 167-82
- Smuts, J. 1959. Die woordeskat van 'n Voortrekker. *TWK (NR)* 19(2): 213-17
- Smuts, J. 1964. Betekenisontwikkeling in Afrikaans. *TVV* 20(4): 22-31
- Smuts, J. 1969. Leksikologiese aantekeninge by die Kaapse Taalgief. *TVV* 25(2): 22-6
- Van Oordt, L.C. 1978. Verwysings na volkskundige gegewens in die Kaapse argief, met enkele opmerkinge (1692–1823). *TVV* 34(2): 35-41.

In addition to the above-mentioned works, the two dialect atlases of Afrikaans (Louw 1959, 1975–79) also contain data which are germane to the history of Afrikaans words. In compiling the following bibliography, I made extensive use of Kettley (1970) and Nienaber (1947–68), both of which provided numerous entries of importance to the subject at hand.

The *Woordeboek van die Afrikaanse Taal* (WAT) is of a hybrid nature, having been initially intended as a synchronic, descriptive dictionary, and under subsequent editors altered to include more or less historical material which can be of interest to the etymologist. As discussed by Botha (1996), the treatment of etymology proper in the WAT has been erratic, though certain volumes contain useful word-historical insights, especially volumes VI–X.

My thanks are due to the Suid-Afrika Huis in Amsterdam for allowing me unrestricted access to their holdings and to Corine de Maier for her great help in locating periodicals, and to the staff of the Special Collections of the J.S. Gericke Library at the University of Stellenbosch, especially Mimi van der Merwe who went to great lengths to locate obscure but necessary articles. Were it not for the Interlibrary Loans at the Wilson Library at the University of Minnesota, there would be no bibliography to publish. I am also grateful to Erika Terblanche at the National Afrikaans Literary Museum and Research Centre in Bloemfontein, Republic of South Africa, as well as Crystal Warren and Malcolm Hacksley from the National English Literary Museum in Grahamstown, Republic of South Africa. Many thanks to Tanja Harteveld from the Bureau of the *Woordeboek van die Afrikaanse Taal* in Stellenbosch. Finally, without the support of Anatoly Liberman, this project would never have been undertaken and seen through.

Though articles on Afrikaans etymology are indispensable and were in need of being brought together and annotated, the majority of the more luminary word histories are in books. These books, along with a few articles are listed below. They all have word lists or are otherwise easy to use. Hesseling (1899, 1923), Du Toit (1905), Bosman (1916), and Scholtz (1963, 1972) all lack word lists, and, being the seminal works that they are, are included in the main body of the bibliography.

Notes

1. Gouws, R.H. and F.A. Poneis. 1992. The Development of Afrikaans and the Lexicographical Tradition. Zgusta, Ladislav (Ed.) *History, Languages, and Lexicographers: 77-104*. Tübingen: Niemeyer.
2. Both terms are understood here to indicate a form of Afrikaans that existed from possibly the beginning of the Colony, but certainly by 1775, to roughly the end of the nineteenth century, that was perceived by its speakers as being Dutch, despite its idiosyncrasies.
3. One famous such instance is Mentzel's oft-quoted observation on the use of *ons* for *wij* amongst the womenfolk of the Cape: 'Die Sprache der Landleute ist so wenig reine Holländische Mundart als die teutschen Bauern reines Teutsch sprechen. Die Mannspersonen nehmen das Maul dabei sehr voll, und das Frauenvolk hat Redensarten angenommen, die zuweilen recht lächerlich sind. Zum Exempel. Man frägt etwan, ob sie keine Bibel haben, so erfolgt die Antwort: *Onz heeft geen Bijbel ...* Wenn man sie aber aldann frägt: Wie viel *Unzen* gehen auf ein Pfund? so werden die schamroth.' (Vol. II chap. 7) Mentzel, O.F. 1785-87. *Vollständige geographische und topographische Beschreibung des afrikanischen Vorgebirges der Guten Hoffnung*. Glogau.
4. Scholtz, J. du P. 1965. *Afrikaans uit die vroeë tyd*: 146-204. Cape Town: Nasou.
5. For a thorough discussion of this literature, see Scholtz, J. du P. 1985. Vroeë taalkundige studies. *TGw* 25(2): 82-93.
6. 'Heeft het zijne belangerijke zijde, om op te merken, hoe het Nederduitsch van het eene gewest van *Nederland* van dat van het andere verschilt, en hoe, door vergelijking, het eene taalgebruik het andere toelicht en opheldert, dan kan het ook niet onbelangerijk zijn, de eigenheden van het Kaapsch-Nederduitsch bijéén gesteld te zien.' (Van der Merwe 1972: 5.)
7. 'Het hoofddoel van de volgende verzameling, gelijk men al dadelijk uit den titel van ons werk kan afleiden, was om het Nederduitsch, voor zoo ver de taal, die in deze Kolonie gesproken wordt, dien naam dragen mag, van deels geheel vreemde, deels verminkte woorden en spreekwijzen te zuiveren, of althans den weg daartoe aan te wijzen.' (Van der Merwe 1972: 5.) Luckily his efforts were for nought, as one inhabitant of Montagu states, 'Jij kan ver mij gloo dat die plat Hollans meer gelees wor onder ons boere as die wat Sankion ver ons wil leer in zijn boekies'. Scholtz, J. du P. 1965. *Die Afrikaner en sy taal 1806-1875*: 180. Cape Town: Nasou.
8. For more on this scholarship, see Loubser, C. De B. 1950. *Die studie van die Afrikaanse woordeskat in die 19de eeu*. Unpublished M.A. Thesis. Cape Town: University of Cape Town.
9. Given how much space these articles would take up, the fact that only some of the words in each study would merit entry here, and that they are organized like dictionaries, Du Plooy 1981 and Leal 1978 were not excerpted and are in the list below of easily consulted works.
10. For a thorough discussion of the Malay aspect of the Afrikaans vocabulary, see Poole, C.H.. 1951. *'n Kritiese beskouing van die Maleise en Portugese woorde in Afrikaans*. Unpublished M.A. Thesis. Cape Town: University of Cape Town. For works on Portuguese words, see Leal (1978) in addition thereto. A further study on Portuguese words in Afrikaans, is that by Rebelo, I.M.F. 1982. *Opmerkings oor Portugese woorde in Afrikaans*. Unpublished B.A. Hons. Thesis. Johannesburg: University of the Witwatersrand.

References

- Boshoff, S.P.E.** 1921. *Volk en taal van Suid-Afrika*. Pretoria: J.H. de Bussy.
- Boshoff, S.P.E.** 1936. *Etimologiese woordeboek van Afrikaans*. Cape Town: Nasionale Pers.
- Boshoff, S.P.E. and G.S. Nienaber.** 1967. *Afrikaanse etimologieë*. Pretoria: Die Suid-Afrikaanse Akademie vir Wetenskap en Kuns.
- Bosman, D.B.** 1923. *Oor die ontstaan van Afrikaans*. Amsterdam: Swets & Zeitlinger.
- Botha, Willem.** 1996. Die WAT en etimologie: Word die kringloop voltooi? *Lexikos* 6: 159-170.
- Du Plooy, Floris D.** 1981. *Moontlike raakpunte tussen Platduits en Afrikaans*. Wetenskaplike bydraes van die PU vir CHO, Reeks A: Geesteswetenskappe, Nr. 41. Potchefstroom: Potchefstroom University for CHE.
- Franken, J.L.M.** 1953a. *Taalhistoriese bydraes*. Amsterdam/Cape Town: A.A. Balkema.
- Franken, J.L.M.** 1953b. *Tentatiewe etimologieë*. Annale van die Universiteit van Stellenbosch, Jaargang 27, Reeks B, No. 1. Stellenbosch: University of Stellenbosch.
- Hesseling, D.C.** 1905. *Het Negerhollandsch der Deense Antillen*. Leyden: Sijthoff.
- Kempen, W.** n.d. *Woordvorming en funksiewisseling in Afrikaans*. Cape Town: Nasionale Boekhandel.
- Kempen, W.** n.d. *Samestelling, afleiding en woordsoortlike meerfunksionaliteit in Afrikaans*. Cape Town: Nasou.
- Kettley, G.M. (Comp.).** 1970. *Suid-Afrikaanse linguïstiese bronnegids*. Raad vir Geesteswetenskaplike Navorsing, Publikasiereeks Nr. 11. Johannesburg: Voortrekkerpers.
- Kloeke, G.G.** 1950. *Herkomst en groei van het Afrikaans*. Leyden: Universitaire Pers.
- Leal, L.** 1978. *Die Portugese leksikale elemente in Afrikaans*. Wetenskaplike bydraes van die PU vir CHO, Reeks A: Geesteswetenskappe, Nr. 28. Potchefstroom: Potchefstroom University for CHE.
- Le Roux, T.H.** 1964. *Die dagboek van Louis Trigardt*. Pretoria: J. L. van Schaik.
- Le Roux, T.H.** 1968. Raakpunte tussen Afrikaans en die Nederlandse volkstaal. Deel II: Woordvorm en woordbetekenis. *TGw* 8: 321-356.
- Links, Tony.** 1989. *So praat ons Namakwalanders*. Cape Town: Tafelberg.
- Louw, S.A.** 1959. *Afrikaanse taalAtlas*. Pretoria: University of Pretoria.
- Louw, S.A. in collaboration with L.C. Eksteen.** 1975-79. *Afrikaanse taalAtlas*. New series 1-5. Pretoria: University of Pretoria.
- Malherbe, D.F.** 1924. *Afrikaanse spreekwoorde en verwante vorme*. Bloemfontein: Nasionale Pers.
- Mansvelt, N.** 1884. *Proeve van een Kaapsch-Hollandsch idioticon met toelichtingen en opmerkingen betreffend land, volk en taal*. Cape Town: Van de Sandt de Villiers. Also in Van der Merwe 1971: 129-222.
- Nienaber, G.S.** 1963. *Hottentots*. Pretoria: J.L. van Schaik.
- Nienaber, P.J. (Comp.).** 1947-68. *Bronnegids by die studie van die Afrikaanse taal en letterkunde*. Six volumes. Further volumes published by the Raad vir Geesteswetenskaplike Navorsing, Pretoria and the Nasionale Afrikaanse Letterkundige Museum en Navorsingsentrum, Bloemfontein.
- Patriot Woordeboek Afrikaans-Engels/Patriot Dictionary Cape Dutch-English*. 1902. Paarl: D.F. du Toit. Republished by H.J.J.M. van der Merwe. Pretoria: J.L. van Schaik.
- Pettman, Charles.** 1913. *Africanderisms*. London: Longmans, Green and Co.

- Rademeyer, J.H.** 1938. *Kleurling-Afrikaans. Die taal van die Griekwas en Rehoboth-Basters*. Amsterdam: Swets & Zeitlinger.
- Raidt, Edith H.** 1994. *Historiese taalkunde. Studies oor die geskiedenis van Afrikaans*. Johannesburg: Witwatersrand University Press.
- Scholtz, J. du P.** 1941. *Uit die geskiedenis van die naamgewing aan plante en diere in Afrikaans*. Cape Town/Bloemfontein/Port Elizabeth: Nasionale Pers.
- Scholtz, J. du P.** 1965. *Afrikaans uit die vroeë tyd: Studies oor die Afrikaanse taal en literêre volkskultuur voor 1875*. Cape Town: Nasou.
- Scholtz, J. du P.** 1972. Boustowwe vir 'n historiese woordeboek van Afrikaans. *Afrikaans-Hollands in die agtiende eeu*: 106-180. Cape Town: Nasou.
- Scholtz, J. du P.** 1975. *Naamgewing aan plante en diere in Afrikaans: Bydrae tot 'n geskiedenis van die Afrikaanse woordeskat*. Second edition. Cape Town: Nasou.
- Scholtz, J. du P.** 1985. Afrikaanse woorde en uitdrukkinge — eiegoed of erfgoed? *TGw* 25: 235-90.
- Silva, Penny (Mng. Ed.), Wendy Dove, Dorothea Mantzel, Colin Muller, Madeleine Wright.** 1996. *A Dictionary of South African English on Historical Principles*. Oxford/New York: Oxford University Press.
- Smith, J.J.** 1962. *Op ons taalakker*. Pretoria: J.L. van Schaik.
- Terblanche, H.J. and J.J. Odendaal.** 1966. *Afrikaanse woordeboek. Verklarend met woordaflerings*. Johannesburg: Afrikaanse Pers-Boekhandel.
- Van der Merwe, H.J.J.M.** 1971. *Vroeë Afrikaanse woordelyste*. Pretoria: J.L. van Schaik.
- Van Oordt, L.C.** 1948-1956. *Die Kaapse Taalargief*. Nrs. 1-4 self-published, 1947-48, nrs. 5-10 in *TWK*, Apr. 1950-Apr. 1956.
- Van Wyk, G.J. (Techn. Ed.), A.E. Cloete, A. Jordaan, H.C. Liebenberg en H.J. Lubbe.** 2003. *Etimologiewoordeboek van Afrikaans*. Stellenbosch: Bureau of the WAT.
- Van Wyk, G.J. (Ed.), A.E. Cloete, H.C. Liebenberg, H.J. Lubbe, A. Prinsloo.** 2007. *Etimologiewoordeboek van Afrikaans. Supplement*. Stellenbosch: Bureau of the WAT.
- Veth, P.J.** 2003. *Uit Oost en West*. Noline van der Sijs (Ed.). Reprint of the 1889 edition. Amsterdam/Antwerpen: Veen.
- Von Wielligh, G.R.** 1925. *Ons geselstaal. 'n Oorsig van gewestelike spraak soos Afrikaans gepraat word*. Pretoria: J.L. van Schaik.

Woordelys / Word List

- aalwyn:** 'n plantsoort / aloe, a kind of plant
Nienaber 1966, 1968a; Scholtz 1972
- aandag:** huisgodsdienst / family prayers
Hesseling 1899
- aapskeloeder:** 'n gemene vent / blackguard
Hesseling 1899, 1923; Schuchardt 1885
- aardig:** onaangenaam / unpleasant
Bouman 1925–26; Lubbe 2004
- abba:** ('n kind) op die rug of skouers dra / piggyback, to carry (a child) on the back or shoulders
Hesseling 1899
- abiekwasgeelhout:** 'n houtsoort / tamarisk, a kind of wood
Boshoff 1950
- afdraand:** skuinste / decline
Liebenberg 1976
- Afrikaner:** 'n Afrikaanssprekende persoon / an Afrikaans-speaking person
Scholtz 1972
- afrokkel:** op 'n slinkse manier verkry / to wheedle out of
De Villiers 1961
- ag:** *uitroep* / oh!, alas!, *exclamation*
Scholtz 1985
- agostou:** 'n plantsoort / a kind of plant
Von Wielligh 1921–22
- agterlosig:** nalatig / careless
Malherbe 1922–23
- aia:** kinderoppasser / nanny
Hesseling 1899
- ajoos:** 'n swamsoort / puff-ball, a kind of fungus
Boshoff 1952
- akkerboom:** eikeboom / oak tree
Scholtz 1985
- akkerwanie:** 'n grassoort / cuscus, a kind of grass
Hesseling 1899
- alikeukel:** 'n eetbare seeslak / periwinkle, an edible sea slug
Smith 1913; Te Winkel 1896
- alla:** *uitroep* / (good) gracious, *exclamation*
Bosman 1916; Du Toit 1905; Smith 1913
- als:** 'n plantsoort / wormwood, a kind of plant
Scholtz 1972
- amper:** byna / almost
Bosman 1916; D'Arbez 1916; Hesseling 1899, 1923; Te Winkel 1896; Van Oordt 1916
- arrie:** *uitroep* / (I) say!, my!, *exclamation*
Bosman 1916; Hesseling 1899; Schuchardt 1885; Smith 1913
- askuus:** ekskuus / excuse me
Scholtz 1972
- aspres:** opsetlik / intentionally
Scholtz 1972

- assegai:** spies / assegai, spear
Bosman 1916; Du Toit 1905; Hesseling 1899, 1923; Te Winkel 1896; Van Oordt 1916
- astrak:** wisselvorm van *ekstrak* / variant of *ekstrak*, extract
Scholtz 1972
- astrant:** parmantig / cheeky, impudent
Scholtz 1972
- atjar:** suurtjies / pickles
Bosman 1916; Hesseling 1899
- baadjie:** 'n bokledingstuk / jacket, an outer garment
Bosman 1916; Hesseling 1899; Malherbe 1922–23; Te Winkel 1896
- baar:** onervare, onbedrewe; ru / inexperienced, unskilled; crude
Bosman 1916; Hesseling 1899, 1923; Nienaber 1954i; Schuchardt 1891; Smith 1938; Te Winkel 1896
- babeljoentjie:** hemelbed / four-poster, canopy bed
Coetzee 1968
- bafta:** 'n growwe blou materiaal / a coarse blue material
Boerevrou; Le Roux 1957
- baie:** erg, veel, heel / very, many, much, quite
Anoniem 1908; Cilliers 1908, 1909; Du Plooy 1966; Hesseling 1899, 1909; Kruyskamp 1946; Scholtz 1963; Smith 1942; Valkhoff 1969; Van Oordt 1916; Von Wielligh 1921–22
- bak:** kos; 'n houer waarin kos opgedien word / food; a container in which food is served
Malherbe 1922–23
- baklei:** vuisslaan; rusie maak / to have a fist-fight; quarrel
Bosman 1916; Hesseling 1899; Scholtz 1972; Te Winkel 1896
- balie:** 'n oop vat / tub, an open vat
Schuchardt 1885
- baljaar:** luidrugtig speel en rondspring / to play, frolic boisterously
Bosman 1916; Hesseling 1897, 1899, 1923; Te Winkel 1896; Valkhoff 1969
- basaar:** 'n Oosterse markplek / bazaar, an Eastern marketplace
Bosman 1916; Hesseling 1899; Van Oordt 1916
- basta:** genoeg! / enough!
Bosman 1916; Hesseling 1899, 1923; Liebenberg 1976; Valkhoff 1969
- beesbyter:** mamba, 'n soort slang / mamba, a kind of snake
Von Wielligh 1921–22
- beghais:** befoeterd / cranky
Bergerson 2004
- beghieks:** mal / crazy
Bergerson 2004
- bejies:** sien *beghieks* / see *beghieks*
- bejieps:** sien *beghieks* / see *beghieks*
- bejoeks:** sien *beghieks* / see *beghieks*
- bejorries:** sien *beghieks* / see *beghieks*
- betjoeks:** sien *beghieks* / see *beghieks*
- betjoins:** sien *beghieks* / see *beghieks*
- betjols:** sien *beghieks* / see *beghieks*
- betjorries:** sien *beghieks* / see *beghieks*
- bermotersanpeer:** 'n peersoort / a kind of pear
Hesseling 1899, 1923; Valkhoff 1955

- biebies:** luise / lice
Vercoullie 1919
- biesroei:** 'n soort riet / a kind of reed
Hesseling 1899, 1923
- bietjie:** 'n klein hoeveelheid / little, a small quantity
Conradie 1994
- biltong:** repe gedroogde vleis / strips of dried meat
Nienaber 1954n
- biskop:** 'n vissoort / musselcracker, a kind of fish
Boshoff 1941
- blaar:** die plat, skyfvormige deel van 'n plant / leaf, the flat, disc-shaped part of a plant
Malherbe 1922-23; Bouman 1922
- blaasbalk:** 'n werktuig om 'n vuur aan te blaas / bellows, an implement to fan a fire
Scholtz 1972
- blatjang:** 'n dik, gekruide vrugtesous / chutney, a thick spiced fruit sauce
Bosman 1937
- bleik:** laat wit word / to bleach
Scholtz 1963
- blus:** weerstandsvermoë / power of resistance
Bouman 1925-26; Malherbe 1922-23
- bobotie:** gebakte kerriemaalvleis / baked curried mince meat
Boerevrou
- boeglam:** afgemat / exhausted
Schuchardt 1885
- Boesman:** 'n lid van 'n Suid-Afrikaanse stam / Bushman, a member of a South African tribe
Den Besten 2004; Lehmann 1945; Nienaber 1952a, 1954f; Van Oordt 1941
- boet:** broer / brother
Bergerson 2002a; Bosman 1916; Du Toit 1905; Hesseling 1923; Vercoullie 1919
- bokman:** mynwerker / miner
De Tollenaere 1957
- bokrin:** 'n roep om bokke in die kraal te jaag / a call to herd goats into the corral
Scholtz 1979; Taalman 1946
- bokruit:** 'n roep om bokke uit die kraal te jaag / a call to herd goats out of the corral
Scholtz 1979; Taalman 1946
- bokveld:** in die uitdrukking "bokveld toe wees", dood wees / in the expression 'bokveld toe wees', to be dead
Lubbe 1967
- bolmakiesie:** met die voete oor die kop / head over heels
Coetzee 1968; Nienaber 1954m
- bonkes:** 'n kort tabakpypie / a short tobacco pipe
Hesseling 1899, 1923
- booi:** 'n manlike bediende / a male servant
Den Besten 2004; Van Oordt 1936
- borrie:** 'n kruiesoort / turmeric, a kind of spice
Hesseling 1899; Smith 1913
- botel:** 'n plek waar bote gehou word / bo(a)tel, a place where boats are kept
Odendal 1964
- bottel:** 'n glashouer met 'n dun nek / bottle
Bosman 1916, 1922-23; Du Toit 1905; Smith 1921-22; Vercoullie 1919

- bra:** eintlik; haas / actually; somewhat
Conradie 1994
- breddie:** gestoofde vleis en groente / stew, stewed meat and vegetables
Hesseling 1899; Smith 1913
- breekspul:** deurmekaarspul / mix-up
Du Plooy 1966
- brei:** ('n vel) bewerk om sag en buigsaam te word / to prepare, tan (a skin) to become soft and pliable
Scholtz 1963
- broesa:** die duiwel / the devil
Hesseling 1899
- burg:** 'n gekastreerde varkmannetjie / barrow, hog, a castrated male pig
Scholtz 1985
- busseltjie:** die vulling agter in 'n rok / bustle, the padding in the back of a dress
Coetzee 1968
- daardie:** dié, ding(e) daar / that, those
Hesseling 1899; Pauwels 1959a; Roberge 2001; Scholtz 1963, 1972
- dalk:** miskien, moontlik / maybe, possibly
Du Plooy 1966; Scholtz 1963, 1972
- dam:** 'n oop opgaarplek vir water / dam, an open storage place for water
Snijman 1977
- die:** *bepaalde lidwoord / the, definite article*
Hesseling 1899, 1919, 1923; Kruisinga 1906; Le Roux 1919; Scholtz 1963, 1972; Van den Berg 1965
- doepa:** toorgoed / magic potion, charms
Du Toit 1905
- doerias:** 'n soort geblomde symateriaal / a kind of floral silk material (vgl./cf. *doria*)
Du Toit 1905
- Doesman:** Duisman / Dutchman
Den Besten 2004
- dogter:** meisie / girl
Du Toit 1905
- dolos:** kneukelbeen, as kinderspeelgoed en vir waarsêery gebruik / knuckle-bone, used as children's playthings and for soothsaying
Boshoff 1951a; Du Toit 1905; Du Toit 1932; Du Toit 1959b; Hesseling 1899; J.B.D. 1932; Lategan 1987; Malherbe 1922–23; Nienaber 1995; Valkhoff 1955
- doodgooi:** sleggebakte brood / badly baked bread
Nienaber 1954k
- doodluiters:** ongeërg, asof onskuldig / unperturbed, playing innocent
Bouman 1928; Du Plooy 1966; Malherbe 1922–23
- dop:** 'n sopie alkoholiese drank / a tot of alcoholic beverage
Du Toit 1958c
- dophou:** in die oog hou / to watch, keep an eye on
Boshoff 1926
- Dopper:** 'n lid van die Gereformeerde Kerk in Suid-Afrika / a member of the Reformed Church in South Africa
Bokhorst 1940; Du Toit 1959c, 1963; H.G.V. 1919
- doria:** 'n soort kledingstof / a kind of fabric (vgl. / cf. *doerias*)
Boerevrou

- drabok:** 'n plantsoort / darnel, a kind of plant
Nienaber 1965, 1968a
- duiwelsterk:** 'n soort kledingstof / a kind of fabric
Smith 1921-22
- dusketyd:** dieselfde tyd / the same time
Labuschagne 1936
- duwejoe:** in die uitdrukking "in die duwejoe wees", onseker wees / in the expression 'in die duwejoe wees', to be unsure
Nienaber 1953, 1954c; Scholtz 1972
- eena, deena:** die eerste woorde van 'n uittelrympie vir kinders / the first words of a children's counting-out rhyme
Nienaber 1954a
- eendjies:** kankerbossie, 'n plantsoort / cancer-bush, a kind of plant
Von Wielligh 1921-22
- eenspaaiërig:** nie lief vir geselskap nie / solitary, not fond of company
Coetzee 1954
- els:** 'n priemvormige werktuig vir die maak van gaatjies / alder, an awl-shaped instrument for making small holes
Scholtz 1972
- emma koelemma:** die eerste reël van 'n liedjie / the first line of a song
Anoniem 1962; Coetzee 1961
- es:** vuurherd / fireplace
Coetzee 1964
- ewwa-trewwa:** 'n plantsoort / a kind of plant
Malherbe 1922-23; Vercoullie 1919
- fieterjasies:** tierlantyntjies / fripperies, knick-knacks
Nienaber 1994a
- filekommis:** in die uitdrukking "iemand filekommis maak", onterf / in the expression 'iemand filekommis maak', to disinherit
Smith 1921-22
- fluit:** 'n entmetode van plante / a grafting method of plants
Coetzee 1968
- froetang:** 'n plantsoort / a kind of plant
Du Toit 1905; Odendal 1966
- gansies:** kankerbossie, 'n plantsoort / cancer-bush, a kind of plant
Von Wielligh 1921-22
- Gatjjeponner:** 'n spotnaam deur Doppers aan lidmate van ander Afrikaanse kerke gegee / nickname given by Doppers to members of other Afrikaans churches
Du Toit 1959c
- gebodder:** die handeling van kwel, pla, lastig wees / the act of vexing, annoying, being troublesome
Du Plooy 1966
- geilsiekte:** 'n veesiekte veroorsaak deur blousuurvergiftiging/ a livestock disease caused by prussic acid poisoning
Smith 1913
- geit:** meisie / girl
Malherbe 1922-23
- geitjie:** 'n kleinerige akkedis met gladde skubbe / gecko, a smallish lizard with smooth scales
Boshoff 1927-28; Du Toit 1905; Smith 1921-22

- gevreet:** gesig / mug, face
Hesseling 1923
- ghantang:** vryer; beminde / suitor; sweetheart
Du Toit 1958a
- ghoen:** 'n groot skietalbaster / taw, a big shooting marble
Hesseling 1899, 1923; Smith 1921–22
- ghwarrie:** gruisgat / quarry, gravel pit
Coetzee 1968; Venter 1968
- glad:** inderdaad / indeed, quite
Conradie 1994
- goël:** deur toertjies bedrieg / to conjure, trick by sleight of hand
Scholtz 1963
- goiing:** grofgeweefde materiaal, dikwels gebruik vir die maak van sakke / hessian, coarsely woven material, often used for making sacks
Bosman 1916; Du Toit 1905; Hesseling 1923; Van Oordt 1916
- graatjie:** stokstertmeerkat, 'n diersoort / suricate, a kind of animal
Boshoff 1927–28; Nienaber 1954d
- gramadoelas:** 'n afgeleë, moeilik bereikbare gebied / outback, a remote, difficultly accessible area
Boshoff 1923–24; Smith 1921–22
- gunter:** doer, ginds / over there, yonder
Scholtz 1972
- gus:** onvrugbaar, steriel / barren, sterile
Smith 1913; Venter 1968
- haai:** *uitroep* / hullo!, hey!, *exclamation*
Bosman 1916; Du Toit 1905
- haai-hoei:** bohaai / ado, hullabaloo
Du Toit 1905
- haasbek:** 'n mond waarin veral die boonste voortande ontbreek / a mouth in which especially the upper front teeth are missing
De Tollenaere 1957
- hake-krukke:** 'n seunspeletjie / a boy's game
Du Plooy 1966
- hakskeen:** die agterste deel van die voet / heel, the back part of the foot
Snijman 1977
- hakskeenpleister:** waardelose papiergeld veral van die vroeëre Zuid-Afrikaanse Republiek / worthless paper money especially from the former Zuid-Afrikaanse Republiek
Boerevrou
- hanepoot:** 'n druifsoort / a variety of grape
Hesseling 1916a
- harslag:** ingewande in die borskas van 'n geslagte dier / pluck, entrails in the thorax of a slaughtered animal
Smith 1913
- hartbeeshuis:** paal-en-klei-huis / wattle and daub hut
Nienaber 1952a, 1952b, 1954e, 1989; Walton 1987
- Hasie Aboel:** 'n kinderspeletjie / catch the hare, a children's game
Kloppers 1951–52, 1952–53a, 1952–53b, 1953
- hê:** infinitiefvorm van *het* / infinitive form of *het*
Scholtz 1972

- hekel:** handwerk vervaardig met gare en 'n haakpen / to crochet
Scholtz 1963
- herneuters:** herneutermes, 'n soort jagmes / Hernehuter, a kind of hunting-knife
Von Wielligh 1921–22
- het:** verboë vorm van *hê* / conjugated form of *hê*
Scholtz 1972, 1985
- hierdie:** ding(e) hier / this, these
Hesseling 1899; Pauwels 1959a; Roberge 2001; Scholtz 1963, 1972
- hoeka:** lank gelede / long ago
Conradie 1994; Hesseling 1899
- hom:** *persoonlike voornaamwoord, voorwerpsvorm, derde persoon enkelvoud, manlik* / him, personal pronoun, object form, third person singular, masculine
Scholtz 1963, 1972
- Horak:** in die uitdrukking "Wat sê Horak?", retoriese vraag om verbasing of twyfel uit te druk / in the expression 'Wat sê Horak?', rhetorical question to express surprise or doubt
Du Toit 1959a; Nienaber 1968b
- hortjie:** 'n raamwerk as bedekking vir 'n deur of venster / louvre, shutter, a covering for a door or window
Lubbe 1970
- hotom:** linksom / to the left
Boerevrou
- Hottentot:** 'n lid van 'n Suid-Afrikaanse stam / a member of a South African tribe
Nienaber 1954g, 1956a, 1956b
- hulle:** *persoonlike voornaamwoord, derde persoon meervoud* / they, personal pronoun, third person plural
Goossens 1994; Scholtz 1963; Vor der Hake 1911
- inspan:** in die uitdrukking "jong osse/kalwers inspan", vomeer / in the expression 'jong osse/kalwers inspan', to vomit
Boshoff 1926
- jaarhonderd:** eeu / century
Hesseling 1899
- jamboes:** 'n boomsoort / rose apple, a kind of tree
Du Toit 1905
- japsnoet:** 'n vrypostige kind / whipper-snapper, an impertinent child
Du Plooy 1966
- jilletjie:** grap / joke
Hesseling 1914
- josie:** die duiwel / the devil
Boshoff 1952; Bosman 1916; Hesseling 1899; Van Oordt 1916
- julle:** *persoonlike voornaamwoord, tweede persoon meervoud* / you, personal pronoun, second person plural
Goossens 1994; Scholtz 1963
- kaaiman:** 'n krokodilagtige reptiel / caiman, a crocodilian reptile
Bosman 1916; Du Toit 1905
- kabaai:** 'n los japon / a loose gown
Bosman 1916; Coetzee 1933; Hesseling 1899
- kaboedel:** 'n klomp / caboodle, a lot, pack
Bouman 1951; Coetzee 1995

- kaduks:** siekerig / unwell
Coetzee 1995
- kafferkoring:** sorghum / sorghum
Boshoff 1941; Laubscher 1941
- kafoefel:** vry / to canoodle
Coetzee 1995
- kaiings:** oorblyfsels wanneer stukkie dierevet uitgebraai word / greaves, remains after bits of animal fat have been rendered
Bosman 1937; Scholtz 1963
- kalant:** in die uitdrukking " 'n ou kalant, lank in die land", 'n ervare, uitgeslape persoon / in the expression ' 'n ou kalant, lank in die land', an experienced, shrewd person
Bouman 1925–26
- kalfater:** 'n skip herstel / caulk, to repair a ship
Coetzee 1968
- kalkoen:** 'n pluimveesoort / turkey, a kind of poultry
Nienaber 1954l
- kalossies:** 'n blomsoort / ixia, a kind of flower
Smith 1913
- kamas:** 'n leerbedekking vir die kuite / legging, gaiter, a leather covering for the calves
Hesseling 1899, 1923
- kamma:** kwansuis / ostensibly, supposedly
Hesseling 1899
- kankerbossie:** 'n plantsoort / cancer-bush, a kind of plant
Von Wielligh 1921–22
- kannie-koenie:** jaloers / jealous
Hesseling 1916a; Pienaar 1928–29
- kaparring:** 'n Maleise houtsandaal / a Malay wooden sandal
Smith 1913
- kapater:** 'n gekastreerde bokram / a castrated he-goat
Bosman 1916; Hesseling 1899, 1914; Smith 1913; Van Oordt 1916
- kapel:** kobra, 'n slangsoort / cobra, a kind of snake (vgl./cf. *pelloe*)
Smith 1921–22
- kapok:** sneeu / snow
Du Toit 1905
- karba:** 'n kan vir vloeistowwe / carboy, a container for liquids
Hesseling 1899
- karbonkel:** 'n edelgesteente / garnet, a precious stone
Coetzee 1995
- kardoes:** papiersak / paper bag
Bosman 1916; Hesseling 1899; Smith 1913
- Kardoes:** in die uitdrukking "bende van Kardoes", 'n rumoerige groep / in the expression 'bende van Kardoes', a rowdy group
Boerevrou
- karet:** 'n bagasierak agterop 'n kar of wa / a luggage carrier at the back of a car or wagon
Odendal 1964; Scholtz 1963, 1985
- karjol:** plesier, pret maak / to make merry, have fun
Coetzee 1995

- karkatjie:** 'n sweertjie aan die ooglid / sty(e), a pustule on the eyelid
De Villiers 1969
- karos:** 'n kombors van diervel / a blanket of animal skin
Hesseling 1899
- kasarm:** 'n reeks skakelhuisies, 'n vervalle gebou / a row of semidetached houses, a ramshackle building
Scholtz 1963
- kastig:** kwansuis / ostensibly, supposedly
Hesseling 1899
- kastrol:** 'n komvormige pot / saucepan
Valkhoff 1955
- katel:** ledekant / bedstead
Bosman 1916; Hesseling 1899; Liebenberg 1976; Schuchardt 1885, 1891; Van Oordt 1916
- katjeepering:** 'n blomsoort / gardenia, a kind of flower
Vercoullie 1919
- katools:** jags, wulps / ruttish, horny
Du Plooy 1966
- kattejag:** in die uitdrukking "op die kattejag uit wees", op 'n vrytog uit wees / in the expression 'op die kattejag uit wees', to be out courting
Du Plooy 1966
- keil:** 'n hoë hoed met 'n silindervormige bol / top hat
De Tollenaere 1957; Hesseling 1923
- kekkelbek:** babbelbek, skinderbek / chatterbox, gossip
Du Plooy 1966
- kennetjie:** 'n seunspeletjie / a boy's game
Du Toit 1962
- keps:** pet / cap
Bergerson 2004
- kerbollie:** baadjie / jacket
Boerevrou
- kerjaker:** baljaar / to frolic
Coetzee 1995; Du Plooy 1966
- kerjoel:** deurmekaar beweeg / to swarm, teem
Coetzee 1995
- kê:** (gekookte) dikmelk / (boiled) sour milk
Scholtz 1963
- kiaat:** 'n houtsoort / Cape teak, a kind of wood
Boshoff 1951b; Hartwig 1973; Smith 1913
- kiepersol:** 'n boomsoort / umbrella-tree, a kind of tree
Bosman 1916; De Bruin 1978; Du Toit 1905; Hesseling 1923
- kierang:** kul by speletjies / to cheat at games
Odendal 1966; Te Winkel 1896
- kiets:** gelykop / quits
Smith 1921–22
- kiewiet:** 'n voëlsoort / plover, peewit, a kind of bird
Scholtz 1972
- klapper:** kokosneut / coconut
Bosman 1916; Du Toit 1905
- klip:** 'n stuk(kie) van 'n gesteente, rots / stone, rock
Hesseling 1899

- kluitjie:** in die uitdrukking "kluitjies bak", leuens vertel / in the expression 'kluitjies bak', to tell untruths
Lubbe 1968
- koedoe:** 'n soort wildsbok / kudu (koodoo), a kind of antelope
Bosman 1916; Du Toit 1905; Smith 1913; Van Oordt 1916
- koël:** 'n bol- of silindervormige projektiel om uit 'n vuurwapen af te skiet / bullet
Scholtz 1963
- kojewel:** 'n vrugtesoort/ guava, a kind of fruit
Bosman 1916; Du Toit 1905; Scholtz 1972; Smith 1913
- Koelie:** Indiër / Coolie, Indian
Bosman 1916
- koerang:** sien *kierang* / see *kierang*
- koernappers:** groot, lomp voete / big, clumsy feet
Coetzee 1968
- koert:** hof / court
Bosman 1916; Hesseling 1899, 1923
- koggelman:** 'n soort akkedis / agama, a kind of lizard
Smith 1913
- komers:** 'n warm bedekking om onder te slaap / blanket
Liebenberg 1976
- kondee:** bolla / chignon
Hesseling 1899
- kool:** in die uitdrukking "by my kool (kolie) waar", dit is werklik waar / in the expression 'by my kool (kolie) waar', it is really true
Du Toit 1905
- kraal:** 'n omheinde ruimte vir vee / corral
Bosman 1916; Hesseling 1899
- krans:** 'n regaf, steil rots / cliff
Odendal 1994
- kras:** liggaamlik en geestelik flink / robust
Te Winkel 1896
- kremetart:** verpoeierde wynsteen / cream of tartar
Hesseling 1914
- Kretie:** in die uitdrukking "Kretie en Pletie", gepeupel / in the expression 'Kretie en Pletie', rabble
Hesseling 1899, 1923
- kriek:** 'n soort insek / cricket, a kind of insect
Bosman 1916; Du Toit 1905
- kuier:** besoek aflê / to visit
Van Ginneken 1927
- laai:** gewoonte / habit
Hesseling 1897, 1899
- laksel:** vermicelli / vermicelli
Hesseling 1899
- laksman:** 'n voëlsoort / fiscal shrike, a kind of bird
Hesseling 1899
- lawaai:** 'n harde geraas / a loud noise
Scholtz 1963
- lat:** wisselvorm van *dat* / variant of *dat*, that
Leys 2005; Paardekooper 1990

- lekseltjie:** 'n klein bietjie / a small quantity
Du Plooy 1966
- lemoen:** 'n vrugtesoort / orange, a kind of fruit
Bosman 1916; Hesseling 1899; Smith 1921–22
- lepaai:** kombers / blanket
Nienaber 1954j
- liefêr:** geneig / inclined
Le Roux 1957
- lieplapper:** niksnuts / good-for-nothing
Du Toit 1905
- likkewaan:** 'n akkedisagtige reptielsoort / leguan, a kind of lacertine reptile
Scholtz 1972
- loesing:** 'n pak slae / spanking
Hesseling 1923; Van Oordt 1936
- lukwart:** 'n vrugtesoort / loquat, a kind of fruit
Bosman 1916; Hesseling 1899
- lumier:** daeraad, oggendskemering / dawn, daybreak
Valkhoff 1955
- lyfstuk:** 'n stuk wat by voorkeur gespeel of gesing word / a favorite song to play or sing
Franken 1938
- ma:** moeder / mother, mum
Bosman 1916; Hesseling 1899, 1923
- maai:** moeder / mother
Hesseling 1897, 1899
- maaidekaai:** 'n baasspelerige persoon / a bossy person
Boerevrou
- maaiër:** die larwe van 'n vlieg of brommer / maggot
Scholtz 1963
- maaifoedie:** derduiwel, skelm / blackguard, scoundrel
Coetzee 1961
- makoppa:** mamba, 'n slangsoort / mamba, a kind of snake
Von Wielligh 1921–22
- makou:** 'n pluimveesoort / Muscovy duck, a kind of poultry
Hesseling 1899, 1923; Smith 1913
- makrappa:** 'n Swart arbeider / a Black worker
Nienaber 1994b, 1995
- makrol:** amandelkoekie / macaroon
Hesseling 1899, 1923
- maltrap:** 'n ligsinnige, onnadenkende persoon / madcap
Du Toit 1958c
- mamparra:** 'n onverfynde persoon / an uncouth person
Nienaber 1954h, 1995
- mandoeka:** padda / frog
Franken 1954
- mandoor:** toesighouer, opsigter / supervisor, overseer
Bosman 1916
- manewales:** kaskenades, kaperjolle / antics, capers
Labuschagne 1936; Scholtz 1972
- manel:** 'n lang, swart bokledingstuk vir mans / frock-coat
Hesseling 1914; Smith 1913

- mannenaar:** 'n groot geraas maak / to make a loud noise
Boerevrou
- marmotjie:** 'n diersoort / guinea-pig, a kind of animal
Coetzee 1965
- maskie:** miskien; nie teenstaande / perhaps; nevertheless
Bosman 1916; Hesseling 1899, 1923; Meyer 1900; Roberge 2002; Schuchardt 1885; Van Oordt 1936
- meester goed:** medisyne / medicine
Smith 1921–22
- menteneer:** beheer; behartig / to control; manage
Waher 2001
- mielie:** 'n graansoort / maize
Bosman 1916; Hesseling 1899; Smith 1913; Te Winkel 1896
- moeg:** uitgeput / tired, fatigued
Scholtz 1963
- moesoeke:** meester, meerdere / master, superior
Du Toit 1905; Hesseling 1899, 1923
- moggel:** 'n soort varswatervis / a kind of freshwater fish
Scholtz 1963
- molskyn:** molvel / moleskin
Boerevrou
- mombakkies:** 'n masker / a mask
Smith 1913
- mos:** immers, inderdaad / after all, indeed
Bergerson 2002b; Conradie 1995; Roberge 2002; Scholtz 1972
- motjie:** 'n (getroude) Moesliemvrou / a (married) Muslim woman
Scholtz 1963
- moveer:** lastig val, pla / to pester, trouble
Bosman 1916; Hesseling 1899, 1923; Smith 1913; Valkhoff 1955; Van Oordt 1916
- nartjie:** 'n vrugtesoort / tangerine, mandarin, a kind of fruit
Franken 1927; Hesseling 1899
- nè:** is dit nie so nie? / isn't that so?
Venter 1968
- nenna:** kinderoppasser / nanny
Boerevrou; Hesseling 1899, 1923
- nimlik:** einste / very same
Scholtz 1979
- nô:** meisie / girl
Hesseling 1899, 1923; Scholtz 1963; Schuchardt 1891; Smith 1913, 1921–22; Valkhoff 1969
- nonna:** jong meesteres / young mistress
Du Toit 1905
- nooi:** sien *nô* / see *nô*
- norring:** 'n groot aantal / a mighty lot
Boerevrou
- nuut:** pas of onlangs vervaardig of bekom / new
Scholtz 1963
- oes:** siek; sleg / sick; bad
Pauwels 1959b
- oes:** ryp gewasse insamel / to harvest
Snijman 1977

- oesbeentjie:** dolos / knuckle-bone
Van Blerk 1958
- oggend:** môre / morning
Scholtz 1963
- olienhout:** 'n houtsoort / wild-olive, a kind of wood
Scholtz 1972
- ongans:** siek / sick
Schoonees 1950
- ons:** *persoonlike voornaamwoord, onderwerpvorm, eerste persoon meervoud / we, personal pronoun, subject form, first person plural*
Bosman 1916; Bouman 1924–25; Goossens 1994; Hesseling 1899, 1916a, 1916b, 1922b, 1923; Kruisinga 1906; Malherbe 1924–25; Meyer 1900; Paardekooper 1969; Scholtz 1963; Schuchardt 1885; Schuhmacher 1973; Smith 1913; Van Oordt 1936; Van Rensburg 1996; Vor der Hake 1911
- ontwil:** in die uitdrukkings "om ... (se) ontwil", "om die ontwil van ...", ter wille van / in the expressions 'om ... (se) ontwil', 'om die ontwil van ...', for the sake of
Scholtz 1972
- oond:** 'n geslote ruimte om in te bak / oven
Scholtz 1972
- oop:** nie toe nie / open, not closed
Scholtz 1972
- oorkrabbetje:** oorbel / earring
Hesseling 1899; Schuchardt 1891
- orlams:** taamlik beskaafd; ervare of bedrewe / reasonably civilized; experienced or skilled
Bosman 1916; Hesseling 1899, 1923; Nienaber 1954i; Schuchardt 1891; Smith 1938; Te Winkel 1896; Van Oordt 1916
- otjie:** vark / pig
De Tollenaere 1957
- ou:** kêrel / fellow, guy
Botha 1976; Le Roux 1936; P. 1911
- ousanna:** vuurstengeweer, pangeweer / flintlock (musket), matchlock (musket)
Hesseling 1923; Vercoullie 1919
- pa:** vader / dad, father
Bosman 1916; Hesseling 1899, 1923
- paai:** aanspreekvorm vir 'n ouer man / form of address for an older man
Bosman 1916; Hesseling 1897, 1899, 1923; Liebenberg 1976
- paiboelie:** iemand met wie kinders bang gemaak word / bogey(man), somebody with whom children are threatened
Du Toit 1905; Smith 1921–22
- paljas:** toorgoed / magic objects
Du Toit 1905
- pamperlant:** flikflooi / to cajole, wheedle
De Tollenaere 1957; Malherbe 1922–23
- papelellekoors:** 'n voorgewende siekte / a feigned illness
Hesseling 1923
- pakkuil:** 'n plantsoort / bulrush, a kind of plant
Smith 1921–22; Von Wielligh 1921–22
- pasella:** present / as a present
Smith 1921–22

- pawieperske:** 'n vrugtesoort / white clingstone peach, a kind of fruit
Hesseling 1899, 1923
- pelloe:** kobra, 'n slangsoort / cobra, a kind of snake (vgl./cf. *kapel*)
Malherbe 1922–23
- perske:** 'n vrugtesoort / peach, a kind of fruit
Smith 1921–22; Von Wielligh 1921–22
- piekel:** met moeite dra / to lug, drag
Du Toit 1905
- pienang:** 'n gereg van gekruide gestoofde vleis / a dish of spiced stewed meat
Du Toit 1905
- piering:** 'n bordjie waarin 'n koppie staan / saucer
Bosman 1916; Valkhoff 1969; Van Oordt 1916
- pikkenien:** 'n jong Swart seun / piccanin, a young Black boy
Hesseling 1899, 1914, 1923; Smith 1921–22; Valkhoff 1955, 1969
- pikkewyn:** 'n voëlsoort / penguin, a kind of bird
Scholtz 1972
- platanna:** 'n paddasoort / clawed toad, a kind of frog
Smith 1913
- poelpetaat:** tarentaal / guinea-fowl
Coetzee 1969
- poerbasedan:** tot tydverdryf / for passing the time
Hesseling 1899; Schuchardt 1885
- pondok:** hut / hut, shack
Bosman 1916; Hesseling 1899; Van Oordt 1916
- portfesiedeur:** 'n dubbele deur wat kan terugvou / a double door which can fold back
Scholtz 1963; Smith 1913
- potjierol:** vetsak / pudge, fatty
De Tollenaere 1957
- prinjeel:** wisselvorm van *prieel* / variant of *prieel*, pergola
Scholtz 1972
- rabbedoe:** 'n wilde, uitgelate kind; 'n seunsagtige meisie / a boisterous, uninhibited child; a tomboy
Bosman 1916; Hesseling 1899, 1923; Schuchardt 1891; Van der Meulen 1917
- ramkie:** 'n eenvoudige soort kitaar / a simple kind of guitar
Scholtz 1963; Smith 1913, 1921–22
- ransel:** 'n pak slae gee / to spank, flog
Lubbe 1968
- rappelkops:** duiselig / dizzy
Hesseling 1899
- rasper:** gereedskap met tandagtige uitsteeksels / rasp, grater, utensil with toothlike extensions
Scholtz 1972
- ratel:** 'n diersoort / badger, a kind of animal
Boshoff 1941
- rêrig:** werklik / really
Nienaber 1991; Scholtz 1963
- resep:** voorskrif vir die bereiding van kos of drank / recipe
Scholtz 1963
- resnawel:** redelik, werklik / reasonably, really
Nienaber 1968a; Waher 2001

- rinneweer:** verniel / to ruin, damage
Scholtz 1972
- rissie:** 'n peul van 'n soort speseryplant / chilli, the pod of a kind of spice plant
Bosman 1916; Hesseling 1899; Smith 1913; Van Oordt 1916
- rondawel:** 'n ronde hut / a round hut
Du Toit 1926–27; Hesseling 1914; Nienaber 1968a
- rooms:** struikelend (ten opsigte van 'n perd) / faltering (with regard to a horse)
De Tollenaere 1957
- rottang:** 'n plantsoort / rattan, cane, a kind of plant
Bosman 1937; Scholtz 1963, 1972
- rysmier:** 'n soort insek / termite, a kind of insect
Nienaber 1954b
- sa:** *uitroep om 'n hond aan te hits / hist!, exclamation to set a dog on*
Smith 1913; Van Oordt 1936
- sambok:** 'n swaar sweep / sjambok, a heavy whip
Hesseling 1899; Te Winkel 1896
- sambreel:** 'n opvoubare son- of reënskerm / umbrella
Bouman 1922; Hesseling 1899; Te Winkel 1896
- sammajoa:** om 't ewe / all the same
Du Toit 1905
- sarlot:** 'n soort ui / shallot, a kind of onion
Scholtz 1972
- se:** *besitlike partikel / possessive particle*
Schuchardt 1885; Scholtz 1963
- seekoei:** 'n diersoort / hippopotamus, a kind of animal
Hesseling 1914; Smith 1913, 1921–22; Von Wielligh 1921–22
- sens:** 'n sekelvormige werktuig om graan mee te maai / scythe
Scholtz 1963, 1972
- seroet:** sigaar / cheroot, cigar
Hesseling 1899
- seties:** 'n soort dans / scottische, a kind of dance
Malherbe 1922–23
- seun:** 'n kind van die manlike geslag / boy
Bosman 1916; Hesseling 1899, 1923
- sieal:** wat teenspoed veroorsaak / which causes misfortune
De Villiers 1974; Malherbe 1922–23
- sielverkoper:** 'n persoon wat matrose onder valse voorwendsels werf / soul-trader, a person who recruits sailors under false pretexts
Hesseling 1916a
- siembamba:** in die liedjie "Siembamba, mama se kindjie" / in the song 'Siembamba, mama se kindjie'
Alberts 1955, 1961, 1989; Boshoff 1917
- sies:** *uitroep van walging / bah, exclamation of disgust*
Bosman 1916; Du Toit 1905
- skaai:** vaslê, steel / to pinch, steal
Smith 1913
- skaaprin:** 'n roep om skape in die kraal te jaag / a call to herd sheep into the corral
Taalman 1946
- skaapruit:** 'n roep om skape uit die kraal te jaag / a call to herd sheep out of the corral
Taalman 1946

- skalk:** skelm / rogue, rascal
Lubbe 1970
- skalt en walt:** rondsletter / to saunter, idle about
Malherbe 1922–23
- skinder:** agteraf kwaadpraat / to slander, gossip
Du Plooy 1966; Scholtz 1972
- skoot:** keer, maal / time
Du Plooy 1966
- smous:** 'n rondreisende handelaar / pedlar, an itinerant trader
Meyer 1900
- snaaks:** vreemd / weird
Coetzee 1968
- soebat:** smeek / to plead
Bosman 1916; Hesseling 1899, 1916b, 1923; Te Winkel 1896; Van Oordt 1916
- soms:** somtyds; af en toe / sometimes; now and then
Scholtz 1963
- soiibrand:** pirosis / heartburn, pyrosis
Scholtz 1963
- sosatie:** blokkies gekruide vleis aan 'n pen gaargemaak / kabob, cubes of spiced meat
cooked on a skewer
Du Toit 1905
- spandabel:** verkwistend / wasteful, extravagant
Waher 2001
- spens:** 'n vertrek waar voedsel gebêre word / pantry
Bosman 1916; Hesseling 1897, 1899, 1914; Liebenberg 1976
- spons:** sponning / bung
Bergerson 2004
- sproei/spru:** ontsteking van die mondslymvlies / thrush, sprue
Scholtz 1963
- stapelgek:** heeltemal gek, mal / stark, raving mad
Lubbe 1968
- stawel:** wisselvorm van *stapel* / variant of *stapel*, stack
Scholtz 1972
- stols:** trots / proud
Hesseling 1899
- stoof:** 'n toestel vir kook of verwarming / stove
Bouman 1922
- strandjut:** 'n diersoort / beach hyena, a kind of animal
Coetzee 1944
- strawasie:** moeilikheid / difficulty
Hesseling 1923
- streepsuiker:** 'n afranseling / a whooping
Malherbe 1922–23
- stryk:** in die uitdrukking "van stryk af", in die war, verbouereerd / in the expression
'van stryk af', in a tangle, flustered
Lubbe 1970
- sulle:** *persoonlike voornaamwoord, derde persoon meervoud / they, personal pronoun, third person plural*
Scholtz 1963

- swael:** 'n voëlsoort / swallow, a kind of bird
Scholtz 1972
- swernoot:** skurk / rascal
Hesseling 1899; Smith 1913
- swets:** vloek / to curse
Lubbe 1968
- swetterjoel:** menigte, klomp / swarm, multitude
Scholtz 1972
- ta:** pa / dad
Bosman 1916; Du Toit 1905
- tabeetjies:** presentjies / small presents
Hesseling 1899, 1923
- talie:** in die uitdrukking "rondomtalie draai" / in the expression 'rondomtalie draai',
turn round and round
Bosman 1916; Hesseling 1899, 1923; Liebenberg 1976
- tamaai:** enorm / enormous
Hesseling 1899; Scholtz 1963; Smith 1913, 1921–22; Valkhoff 1969
- tamboekie:** Temboe / Tembu
Scholtz 1972
- tameletjie:** 'n soort lekkergoed / a kind of candy
Scholtz 1972
- tarentaal:** 'n voëlsoort / guinea fowl, a kind of bird
Coetzee 1969; Nienaber 1954; Scholtz 1972; Smith 1921–22
- tasal:** repies gesoute en gekruide vleis wat winddroog gemaak is / strips of salted and
spiced meat which have been air-dried
Du Toit 1905; Smith 1913
- teak:** djatihout, 'n houtsoort / teak, a kind of wood
Boshoff 1951b; Hartwig 1973
- teiken:** skyf; doelwit / target; goal
Scholtz 1963
- ter duiwel:** *uitroep, vloekwoord* / the deuce!, *exclamation, swear word*
Hesseling 1899
- tet:** vrouebors / breast
Du Toit 1905
- tiekie:** trippens / tickey, three pence
Jeffreys 1952; Smith 1921–22
- tier:** luiperd, 'n diersoort / leopard, a kind of animal
Bouman 1928; Hesseling 1899
- tjalie:** sjaal / shawl
Du Toit 1905
- tjap:** stempel / stamp
Bosman 1916; Hesseling 1899
- tjôwa:** probeer / to try
Scholtz 1972
- toering:** 'n Maleise strooihoed / a Malay straw hat
Du Toit 1905
- tof:** deftig, waardig / smart, dignified
Rautenbach 1978
- toiings:** flenters, pluiings / rags, tatters
Bosman 1937; Malherbe 1922–23; Scholtz 1963

- trassie:** 'n onvrugbare dier / a barren animal
Bosman 1916; Du Toit 1905
- trawaat:** 'n sterk wind / a strong wind
Hesseling 1899, 1923
- trawal:** moeilikheid / trouble
Hesseling 1897, 1899
- tronk:** gevangenis / prison
Hesseling 1897, 1899, 1922a, 1923; Valkhoff 1955, 1969
- twak:** tabak / tobacco
Scholtz 1972
- tweealfaamsvat:** 'n soort vat / a kind of vat
Le Roux 1957
- u:** *persoonlike voornaamwoord, tweede persoon hoflik / personal pronoun, second person polite*
Scholtz 1963; Vor der Hake 1911
- uitwiks:** 'n pak slae gee / to spank
Hesseling 1899
- vaderlands:** met betrekking tot Nederland / referring to the Netherlands
Snijman 1977
- vark:** 'n diersoort / pig, a kind of animal
Scholtz 1972
- veels geluk:** baie geluk / congratulations
Hesseling 1899; Schuchardt 1885
- vergange:** vervloë; onlangs / bygone; recently
Hesseling 1899
- vergeet-my-nietjie:** 'n blomsoort / forget-me-not, a kind of flower
Anoniem 1921
- vieruurtjie:** 'n blomsoort / afternoon-lady, a kind of flower
Bosman 1916; Du Toit 1905; Franken 1954; Hesseling 1923; Vercoullie 1919
- viets:** uitgedos, uitgevat / dressed up, turned out
Hesseling 1899; Snijman 1977
- vlei:** 'n moerassige laagte / a marsh
Scholtz 1972
- voetsek:** gaan weg! / scam!
Du Plooy 1966; Schuchardt 1885; Smith 1921–22
- volstruis:** 'n voëlsoort / ostrich, a kind of bird
Bosman 1916; Du Toit 1905; Smith 1913, 1921–22; Vercoullie 1919
- vreemd:** wisselvorm van *vreemd* / variant of *vreemd*, strange
Scholtz 1963, 1985
- waai:** gaan / to go
Hesseling 1897
- waster:** sluitring / washer
Scholtz 1972
- watter:** *vraende voornaamwoord* / what, which, *interrogative pronoun*
Bosman 1916; Hesseling 1899, 1923; Meyer 1900; Von Wielligh 1921–22
- weebelaar:** 'n plantsoort / plantain, a kind of plant
Smith 1913
- werda:** wie's daar? / who goes there?
Van Oordt 1936
- ystervark:** 'n diersoort / porcupine, a kind of animal
Brill 1914

Bibliografie / Bibliography

- Alberts, Andries.** 1955. Siembamba. *TVV* 12(1-2): 55-6.
siembamba
- Alberts, Andries.** 1961. Oorsprong van siembamba. *Standpunte* 14(3): 62-5.
siembamba
- Alberts, Andries.** 1989. Siembamba. *TVV* 45(1): 23-5.
siembamba
- Anoniem.** 1908. Het woord *banje*. *Volkstem* 8 December: 4.
baie
- Anoniem.** 1921. Die blomnaam *vergeet-my-nietjie*. *Huisgenoot* September: 178.
vergeet-my-nietjie
- Anoniem.** 1962. Aanvulling: *emma koelemma*. *TVV* 18(1): 33.
emma koelemma
- Bergerson, Jeremy.** 2002a. The Etymology of Afrikaans *boet* and English *buddy*, *Boots*. *LB* 91: 63-71.
boet
- Bergerson, Jeremy.** 2002b. An Etymology of Afrikaans *mos*. *TNTL* 118: 335-41.
mos
- Bergerson, Jeremy.** 2004. Emphatic -s in Modern Germanic. *IJGLSA* 9: 83-100.
beghais: 90-4; beghiaks: 91-2; bejies: 92; bejieps: 92; bejoeks: 91; bejorries: 91; betjoeks: 91; betjoins: 91; betjols: 91; betjorries: 91; keps: 88; spons: 88
- Boerevrou.** 1926. Die Boerevrou se woordelys. April-December.
April nenna: 5; May doria: 5; hotom: 5; nenna: 4, 5; June bafta: 5; bobotie: 5; kerbollie: 5; molskyn: 5; August hakskeenpleister: 4; hotom: 6; maaidekaai: 5; October mannewaar: 5; norring: 5; December Kardoos: 63
- Bokhorst, M.** 1940. Die woord *dopper*. *HS* 2: 71-5.
Dopper
- Boshoff, S.P.E.** 1917. Siembamba. *Brandwag* 25 March: 326.
siembamba
- Boshoff, S.P.E.** 1923-24. Oor die afleiding van *g(r)amadoelas* of *kamadoelas*. *TWK* 2: 15-23.
gramadoelas
- Boshoff, S.P.E.** 1926. Review: D.F. Malherbe, *Afrikaanse spreekwoorde en verwante vorme*. *TWK* 5: 28-32.
dophou: 30-1; inspan 31-2
- Boshoff, S.P.E.** 1927-28. Boustowwe vir die geskiedenis van Afrikaans. Latrobe (1816). *TWK* 6: 183-94.
geitjie: 191-2; graatjie: 192
- Boshoff, S.P.E.** 1941. Review: J. du P. Scholtz, *Uit die geskiedenis van die naamgewing aan plante en diere in Afrikaans*. *Huisgenoot* 6 June: 33, 35, 63.
biskop: 33, 35; kafferkoring: 35; ratel: 35
- Boshoff, S.P.E.** 1950. Abiekwasgeelhout. 'n Etimologiese bydrae. *TWK (NR)* 10(1): 89-91.
abiekwasgeelhout
- Boshoff, S.P.E.** 1951a. Divagasies oor *dolosse*. *TWK (NR)* 11(1): 81-94.
dolos
- Boshoff, S.P.E.** 1951b. *Kiaat* en *teak*. *TWK (NR)* 11(1): 95-8.
kiaat; teak
- Boshoff, S.P.E.** 1952. *Ajoos*. *TWK (NR)* 12(1): 102-3.
ajoo; josie: 102

Bosman, D.B. 1916. *Afrikaans en Maleis-Portugees*. Groningen: P. Noordhoff.

alla: 49-50; *amper*: 50; *arrie*: 50; *assegaai*: 50-1; *atjar*: 51; *baadjie*: 51; *baar*: 51-2; *baklei*: 52-3; *baljaar*: 53; *basaar*: 53-4; *basta*: 81; *boet*: 54-5; *bottel*: 55; *goiing*: 56; *haai*: 56; *josie*: 56-7; *kaaiman*: 57; *kabaai*: 57; *kapater*: 57-8; *kardoes*: 58; *katel*: 58; *kiepersol*: 58-9; *klapper*: 59; *koedoe*: 59; *koejawel*: 60; *Koelie*: 60; *koert*: 60; *kraal*: 61; *kriek*: 61; *lemoen*: 61-2; *lukwart*: 62; *ma*: 75-6; *mandoor*: 62; *maskie*: 62; *mielie*: 62-3; *moveer*: 63; *ons*: 101; *oorlams*: 63-4; *pa*: 75-6; *paai*: 64-5; *piering*: 65; *pondok*: 65-6; *rabbedoe*: 66; *rissie*: 66-7; *seun*: 73-4; *sies*: 67; *soebat*: 68; *spens*: 55-6; *ta*: 68-9; *talie*: 69; *tjap*: 69; *trassie*: 69-70; *vieruurtjie*: 70; *volstruis*: 70-1; *watter*: 101-2

Bosman, D.B. 1922-23. Oor die afleiding van *bottel*. *TWK* 1: 67-71.

bottel

Bosman, D.B. 1937. 'n Ondersoek na die gevelariseerde *-ing* in Afrikaans. *NT* 31: 58-70.

blatjang: 64; *kaiings*: 69; *rottang*: 64; *toiings*: 69

Botha, J.P. 1976. Die prenominale, unieke *ou*. *FS vdMerwe*: 21-5.

ou

Bouman, A.C. 1922. Review: S.P.E. Boshoff, *Volk en taal van Suid-Afrika*. *Huisgenoot* March 1922: 469-70.

blaar: 469; *sambreel*: 469; *stoof*: 469

Bouman, A.C. 1924-25. Afrikaans *ons* naas Nederlands *wij*. *TWK* 3: 161-2.

ons

Bouman, A.C. 1925-26. Review: D.F. Malherbe, *Afrikaanse spreekwoorde en verwante vorme*. *TWK* 4: 92-8.

aardig: 95; *blus*: 95-6; *kalant*: 97

Bouman, A.C. 1928. Ontlening en relikten in Afrikaans. *NT* 22: 37-41.

doodluiters: 40; *tier*: 41

Bouman, A.C. 1951. Het voorvoegsel *ka-*. *TNTL* 68: 97-128.

kaboedel: 115

Brill, J. 1914. Die IJstervark. Over een Afrikaanse diernaam en een Oud-Grieks volksgeloof. *Brandwag* 1 June: 1-2, 4.

ystervark

Cilliers, C.A. 1908. Het woord *baing*. *Volkstem* 15 December: 10.

baie

Cilliers, C.A. 1909. 't Woord *banje*. *Volkstem* 9 February: 4.

baie

Coetzee, Abel. 1944. Strandjut. *TVV* 1(2): 51-2.

strandjut

Coetzee, Abel. 1954. Van eensgesind tot dwarstrekkerig. *TVV* 11(2): 1-5.

eenspaaiërig

Coetzee, Abel. 1961. Emma koelemma. *TVV* 17(2): 28-31.

emma koelemma; *maaifoedie*: 28

Coetzee, Abel. 1964. Die es. *TVV* 20(3): 33-7.

es

Coetzee, Abel. 1965. Die *marmotjie*-kaart. *TO*: 89-100.

marmotjie

Coetzee, Abel. 1968. 'n Etimologiese woordeboek vir Afrikaans. Review: Boshoff and Nienaber, *Afrikaanse etimologieë*. *TVV* 24(1): 32-6.

babbelloentjie: 34; *bolmakiesie*: 34; *busseltjie*: 35; *fluit*: 35; *ghwarrie*: 35; *kalfater*: 35; *koernappers*: 36; *snaaks*: 35

- Coetzee, Abel.** 1969. Dinges. *TVV* 25(1): 1-8.
poelpetaat: 1; *tarentaal*: 1
- Coetzee, A.J.** 1933. Die wonderbaarlike swerftogte van *kabaai*. *Vaderland* 1 Desember: 13.
kabaai
- Coetzee, Anna E.** 1983. 'n Vreemde prefiks. *FS Nienaber*: 95-8.
kafoefel
- Coetzee, Anna E.** 1995. *Kaboems, kabolder, kerjaker, karbonkel, karfoefel*: vanwaar die hele kaboedel? *SATT Supp* 28: 27-44.
kaboedel: 33-4; *kaduks*: 32; *kafoefel*: 34-8; *karbonkel*: 32-3; *karjol*: 32; *kerjaker*: 32; *kerjoel*: 32
- Conradie, C.J.** 1994. Van hoeveel tot houding: oor die subjektivering van enkele Afrikaanse kwantore. *FS Raidt*: 49-61.
bietjie: 50-2; *bra*: 52-3; *glad*: 53-4; *hoeka*: 55-6
- Conradie, C.J.** 1995. Die partikel *mos*: 'n semantiese verkenning. *SATT Supp* 28: 45-55.
mos
- D'Arbez.** 1916. 'n Historiese verklaring van 'n Afrikaans woord. *Brandwag* 1 August: 96.
amper
- De Bruin, P.S.** 1978. Die sambreel van die kiepersol. *TVV* 34(2): 7-8.
kiepersol
- Den Besten, Hans.** 2004. The Pidgin (and English) Roots of Afrikaans *booi* and *Boesman*. *FS vSchalkwyk*: 183-94.
Boesman: 188-90; *booi*: 184-88; *Doesman*: 184
- De Tollenaere, F.** 1957. Review: J.L.M. Franken, *Tentatiewe etimologieë*. *TNTL* 75: 157-9.
bokman: 158; *haasbek*: 158; *keil*: 158; *otjie*: 159; *pamperlant*: 158; *potjierol*: 159; *rooms*: 159
- De Villiers, Meyer.** 1961. Oor *afrokkel* en *afroggel*. *TGW* 1: 214-6.
afrokkel
- De Villiers, Meyer.** 1969. Die herkoms van *karkatjie*. *TGW* 9: 81-3.
karkatjie
- De Villiers, Meyer.** 1974. Grepe uit die Afrikaanse woordeskat. *FS Kempen*: 31-7.
siaal: 37
- Du Plooy, Floris D.** 1966. Einige Beobachtungen zu niederdeutschen Einschlägen im Afrikaans. *Acta Germanica* 1: 159-66.
baie: 166; *breekspul*: 161; *dalk*: 161; *doodluiters*: 161-2; *gebodder*: 162; *hake-krukke*: 162; *japsnoet*: 162-3; *katools*: 163; *katteljag*: 163; *kekkelbek*: 163; *kerjaker*: 164; *lekseltjie*: 164; *skinder*: 161; *skoot*: 165; *voetsek*: 165
- Du Toit, Heinrich S.** 1932. Oorsprong van die woord *dolos*. *Huisgenoot* 15 April: 39.
dolos
- Du Toit, P.J.** 1905. *Afrikaansche studies*. Ghent: A. Siffer.
alla: 40; *assegaai*: 40-1; *boet*: 41-3; *bottel*: 43-4; *doepa*: 44-5; *doerias*: 45; *dogter*: 69-70; *dolos*: 22-3; *froetang*: 47; *geitjie*: 22; *goiing*: 47-8; *haai*: 48; *haai-hoei*: 48; *jamboes*: 49; *kaaiman*: 49; *kapok*: 70-1; *kiepersol*: 49-50; *klapper*: 50; *koedoe*: 51; *koejawel*: 51-2; *kool*: 52; *kriek*: 52-3; *lieplapper*: 53; *moesoek*: 54; *nonna*: 54-5; *paaiboelie*: 55; *paljas*: 56; *piekel*: 56; *pienang*: 56-7; *sammajoa*: 57; *sies*: 57; *sosatie*: 57-8; *ta*: 58-9; *tasal*: 59; *tet*: 59; *tjalie*: 59-60; *toering*: 60; *trassie*: 60-2; *vieruurtjie*: 45; *volstruis*: 46-7
- Du Toit, S.J.** 1926-27. Die woord *rondawel*. *TWK* 5: 127-8.
rondawel
- Du Toit, S.J.** 1958a. Die malles en die dolles. *Huisgenoot* 4 April: 53.
maltrap
- Du Toit, S.J.** 1958b. Ghantang, die vlugvoetige. *Huisgenoot* 2 May: 59.
ghantang

- Du Toit, S.J.** 1958c. Dis alles doppe. *Huisgenoot* 8 August: 69.
dop
- Du Toit, S.J.** 1959a. Wat sê Horak? *Huisgenoot* 10 April: 35.
Horak
- Du Toit, S.J.** 1959b. Is 'n dolos dol? *TVV* 15(2): 9-15.
dolos
- Du Toit, S.J.** 1959c. Dop, Dopper en Gatjieponner. *TVV* 15(3): 1-14.
Dopper; Gatjieponner: 13-4
- Du Toit, S.J.** 1962. Kennetjie. *TVV* 18(2): 19-22.
kennetjie
- Du Toit, S.J.** 1963. Meurant en die Doppers. *TVV* 19(3): 7-9.
Dopper
- Franken, J.L.M.** 1927. Mededelinge en opinies oor Afrikaans in die neëntiende eeu. *Huisgenoot* 1 July: 41, 43, 57.
nartjie: 43 n. 8
- Franken, J.L.M.** 1938. Review: J.H. Rademeyer, *Kleurling-Afrikaans. Die taal van die Griekwas en Rehoboth-Basters*. *Huisgenoot* 6 May: 45, 65.
lyfstuk: 45
- Franken, J.L.M.** 1954. Review: G.S. Nienaber, *Oor Afrikaans II*. *OEB* 20: 134-6.
mandoeka: 135; vieruurtjie: 135
- Goossens, Jan.** 1994. Hulle. *FS Raidt*: 62-9.
hulle; julle: 62; ons: 66
- Hartwig, G.L.F.** 1973. Klaat en teak. *TGw* 13: 138-42.
klaat; teak
- Hesseling, D.C.** 1897. Het Hollandsch in Zuid-Afrika. *VOAT*: 139-151.
baljaar: 148; laai: 148; maai: 148; paai: 148; spens: 148; trawal: 148; tronk: 148; waai: 149
- Hesseling, D.C.** 1899. *Het Afrikaansch. Bijdrage tot de geschiedenis der Nederlandsche taal in Zuid-Afrika*. Leyden: E.J. Brill.
aandag: 82-3; aapskeloeder: 82; abba: 80; aia: 88; akkerwanie: 88; amper: 88-9; arrie: 89; assegaai: 80; atjar: 89; baadjie: 89-90; baar: 90; baie: 90-1; baklei: 90; baljaar: 90; basaar: 91; basta: 114n; bermotser-sanpeer: 81-2; biesroei: 79; bonkes: 105-6; borrie: 91; bredie: 91; broesa: 91; daardie: 139-40; die: 136-7; dolos: 80; ghoen: 80; hierdie: 139-40; hoeka: 80; jaarhonderd: 83; josie: 92-3; kabaai: 93; kamas: 82; kamma: 81; kapater: 93; karba: 93; kardoos: 93; karos: 81; kastig: 81; katel: 94; klip: 100n; koert: 94; kondee: 107; kraal: 94-5; Kretie: 83; laai: 95-6; laksel: 96; laksman: 96; lemoen: 96; lukwart: 96; ma: 111-2; maai: 96; makou: 96; makrol: 82; maskie: 97; mielie: 97; moesoek: 97; moveer: 97; nartjie: 97; nenna: 97; nôi: 97-8, 107; ons: 137-9; oorkrabbetje: 95; oorlams: 98-9; pa: 111-2; paai: 99-100; pawieperske: 82; pikkenien: 100-1; poerbasledan: 86-7; pondok: 101; rabbedoe: 101; rappelkops: 83; rissie: 101; sambok: 102; sambreel: 102; seroet: 102; seun: 111; soebat: 102; spens: 91-2; stols: 83; swernoot: 84; tabeetjies: 107-8; talie: 102-3; tamaai: 103; ter duiwel: 84; tier: 104-5; tjap: 103; trawaat: 108; trawal: 103; tronk: 103-4; uitwiks: 84; veels geluk: 86; vergange: 87; viets: 82; watter: 140
- Hesseling, D.C.** 1909. 't Woord banje. *Volkstem* 9 February: 4.
baie
- Hesseling, D.C.** 1914. Nieuwe geschriften over het Afrikaans. *Brandwag* 15 January: 501-6.
jilletjie: 501; kapater: 501; kremetart: 501; manel: 501 pikkenien: 501; rondawel: 501; seekoei: 502; spens: 501
- Hesseling, D.C.** 1916a. *Africana II*. De taalkundige opmerkingen van O.F. Mentzel. *TNTL* 35: 275-88.
hanepoot: 276; kannie-koenie: 284; ons: 282; sielverkoper: 277-82

- Hesseling, D.C.** 1916b. Een Afrikaanse dissertatie. Review: D.B. Bosman, *Afrikaans en Maleis-Portugees*. NT 10: 247-56.
ons: 255-6; *soebat*: 255
- Hesseling, D.C.** 1919. Nog eens *die* als lidwoord. NT 13: 95-6.
die
- Hesseling, D.C.** 1922a. De Afrikaners en hun taal. NT 16: 77-83.
tronk: 81
- Hesseling, D.C.** 1922b. Africana III. Het Nederlands aan de Kaap tijdens Van Riebeeck. TNTL 41: 237-45.
ons: 237-8
- Hesseling, D.C.** 1923. *Het Afrikaans. Bijdrage tot de geschiedenis der Nederlandsche taal in Zuid-Afrika*. Leyden: E.J. Brill.
aapskeloeder: 79; *amper*: 82; *assegaai*: 75n, 82; *baar*: 74n, 82; *baljaar*: 83; *basta*: 97n; *bermotsersanpeer*: 77; *biesroei*: 76n; *boet*: 83; *bonkes*: 89-90; *die*: 116-8; *gevreet*: 80; *ghoen*: 77n; *goiing*: 83-4; *kamas*: 77; *keil*: 76n; *kiepersol*: 84, 90; *koert*: 84; *Kretie*: 79; *loesing*: 76n; *ma*: 95-6; *makou*: 85; *makrol*: 77; *maskie*: 47; *moesoek*: 85; *moveer*: 34, 85; *nenna*: 86; *nôi*: 86; *ons*: 119-21; *oorlams*: 86; *ousanna*: 76n; *pa*: 95-6; *paai*: 86; *papelelekoors*: 76n; *pawieperske*: 77; *pikkenien*: 86; *rabbedoe*: 76n; *seun*: 95; *soebat*: 87; *strawasie*: 79; *tabeetjies*: 91; *talie*: 87; *trawaat*: 91-2; *tronk*: 72n, 88; *vieruurtjie*: 83; *watter*: 121-22
- H.G.V.** 1919. Review: Dr. Okulis, *Doppers*. *Huisgenoot* January: 612.
Dopper
- J.B.D.** 1932. Oorsprong van die woord *dolos*. *Huisgenoot* 18 March: 31, 51.
dolos
- Jeffreys, M.D.W.** 1952. Tickey: Origin of the Word? *AN&N* 10: 25-27.
tiekie
- Kloppers, M.H.O.** 1951-52. Hasie Aboel. *TVV* 8(3-4): 117-50.
Hasie Aboel
- Kloppers, M.H.O.** 1952-53a. Hasie Aboel (vervolg). *TVV* 9(1): 26-32.
Hasie Aboel
- Kloppers, M.H.O.** 1952-53b. Hasie Aboel (slot). *TVV* 9(2): 15-9.
Hasie Aboel
- Kloppers, M.H.O.** 1953. Hasie-aboel: verantwoording. *TVV* 10(2): 33-4.
Hasie Aboel
- Kruisinga, E.** 1906. De oorsprong van het Afrikaans. *T&L* 16: 417-39.
die: 420-1; *ons*: 421
- Kruyskamp, C.** 1946. *Banjer, banjerheer, banjaard*. TNTL 64: 51-4.
baie
- Labuschagne, F.J.** 1936. Review: S.P.E. Boshoff, *Etimologiese woordeboek van Afrikaans*. *Koers* 4(3): 32-3.
dusketyd: 33; *manewales*: 33
- Lategan, Felix V.** 1987. Dolosse inspan. 'n Afrikaanse boerekinderspel. *TVV* 43(1): 4-14.
dolos
- Laubscher, F.X.** 1941. Die naam *kafferkoring*. *Huisgenoot* 4 Julie: 45.
kafferkoring
- Lehmann, F. Rudolph.** 1945. 'n Paar gedagtes oor die naam *Boesman*. *TVV* 2(1): 1-6.
Boesman
- Le Roux, J.J.** 1919. Het lidwoord *die*. NT 13: 45.
die

- Le Roux, J.J.** 1936. Die woordjie *ou*. *Huisgenoot* 18 September: 43.
ou
- Le Roux, T.H.** 1957. Taaloudhede. *TVV* 13(3): 1-7.
bafta: 5-6; *liefêr*: 4-5; *tweealfaamsvat*: 3-4
- Leys, Odo.** 2005. Versterkte voegwoorde in het Vlaams: *lat*, *lof* en *nof*. *TT* 57: 111-25.
lat: 111-8
- Liebenberg, H.C.** 1976. Matroostaal in Afrikaans. *Klasgids* 11(3): 5-16.
afdraand: 6; *basta*: 5; *katel*: 6; *komers*: 6; *paai*: 6; *spens*: 6; *talie*: 6
- Lubbe, H.J.** 1967. Opmerkings n.a.v. *hy is bokveld toe*. *TVV* 23(1-2): 20-4.
bokveld
- Lubbe, H.J.** 1968. Opmerkinge n.a.v. enkele etimologiese afleidinge deur proff. Boshoff en Nienaber in *Afrikaanse etimologieë*. *TVV* 24(3-4): 55-8.
kluitjie: 55; *ransel*: 58; *stapelgek*: 55-7; *swets*: 57
- Lubbe, H.J.** 1970. Review: Johan Combrink, *Waar kom daardie woord vandaan?* *TVV* 26(2): 28-9.
hortjie: 29; *skalk*: 29; *stryk*: 29
- Lubbe, Johan.** 2004. *Aardig* in Afrikaans en Nederlands — 'n diachroniese beskrywing van die betekenisverskille. *FS vSchalkwyk*: 255-67.
aardig
- Malherbe, D.F.** 1922-23. Review: S.P.E. Boshoff, *Volk en taal van Suid-Afrika*. *TWK* 1: 112-23.
agterlosig: 121; *baadjie*: 119; *bak*: 115-6; *blaar*: 118-9; *blus*: 122; *dolos*: 122; *doodluiters*: 120; *ewwa-trewwa*: 120; *geit*: 118; *pamperlang*: 122; *pelloe*: 119; *seties*: 123; *sieal*: 122; *skalt en walt*: 120-1; *streepsuiker*: 119; *toings*: 119
- Malherbe, D.F.** 1924-25. Opmerking. *TWK* 2: 162.
ons
- Meyer, Heinrich.** 1900. Der Ursprung der Burensprache. *FS HG*: 83-121.
maskie: 108; *ons*: 112; *smous*: 90; *watter*: 89
- Nienaber, G.S.** 1952a. Twee Afrikaanse woorde. *TWK (NR)* 12(2): 18-25.
Boesman: 18-23; *hartbeeshuis*: 23-5
- Nienaber, G.S.** 1952b. Herkoms van *hartbeeshuisie*. *Vaderland Ekstra* 20 September: 10-11.
hartbeeshuis
- Nienaber, G.S.** 1953. Die duwejoe in wees vir iemand. *Vaderland Ekstra* 11 April: 10-11.
duwejoe
- Nienaber, G.S.** 1954a. 'n Aantekening by *eena*, *deena*. *TVV* 11(1): 15-6.
eena, *deena*
- Nienaber, G.S.** 1954b. Waar kry jy daardie woord? 1. Rysmiere. *Huisgenoot* 9 July: 11.
rysmier
- Nienaber, G.S.** 1954c. Waar kry jy daardie woord? 2. In die dubbeljoe. *Huisgenoot* 16 July: 17.
duwejoe
- Nienaber, G.S.** 1954d. Waar kry jy daardie woord? 5. Graatjiemeerkat. *Huisgenoot* 6 August: 29.
graatjie
- Nienaber, G.S.** 1954e. Waar kry jy daardie woord? 6. Hartbeeshuisie. *Huisgenoot* 13 August: 17.
hartbeeshuis
- Nienaber, G.S.** 1954f. Waar kry jy daardie woord? 7. Boesman. *Huisgenoot* 20 August: 11.
Boesman
- Nienaber, G.S.** 1954g. Waar kry jy daardie woord? 8. Hottentot. *Huisgenoot* 27 August: 23.
Hottentot
- Nienaber, G.S.** 1954h. Waar kry jy daardie woord? 9. Mamparra. *Huisgenoot* 3 September: 35.
mamparra

- Nienaber, G.S.** 1954i. Waar kry jy daardie woord? 10. Hy is nog baar. *Huisgenoot* 10 September: 43.
baar; oorlams
- Nienaber, G.S.** 1954j. Waar kry jy daardie woord? 11. Woorde in Bantoe bewaar. *Huisgenoot* 17 September: 29, 31.
lepaai
- Nienaber, G.S.** 1954k. Waar kry jy daardie woord? 12. Handkous en doodgooi. *Huisgenoot* 24 September: 27, 29.
doodgooi
- Nienaber, G.S.** 1954l. Waar kry jy daardie woord? 13. Kalkoene en tarentale. *Huisgenoot* 8 Oktober: 45, 47.
kalkoen: 45; tarentaal: 45, 47
- Nienaber, G.S.** 1954m. Waar kry jy daardie woord? 15. Bolmakiesie slaan. *Huisgenoot* 29 Oktober: 12-3.
bolmakiesie
- Nienaber, G.S.** 1954n. Waar kry jy daardie woord? 16. Biltong. *Huisgenoot* 5 November: 40-1.
biltong
- Nienaber, G.S.** 1956a. Oor die herkoms van die woord *Hottentot*. *Standpunte* 10(5): 28-46.
Hottentot
- Nienaber, G.S.** 1956b. Oor die herkoms van die woord *Hottentot* (vervolg). *Standpunte* 10(6): 44-55.
Hottentot
- Nienaber, G.S.** 1965. Die bok in *drabok*. *Taalfasette* 1: 54-9.
drabok
- Nienaber, G.S.** 1966. Die wyn in *aalwyn*. *FS Louw*: 108-17.
aalwyn
- Nienaber, G.S.** 1968a. Rondawel. *TVV* 24(3-4): 1-10.
aalwyn: 8; drabok: 8; resnawel: 10; rondawel
- Nienaber, G.S.** 1968b. Wat sê Horak? *Standpunte* 21(5): 31-6.
Horak
- Nienaber, G.S.** 1989. Sonde met *hartbees* in *hartbeeshuis*. *TGw* 29: 180-92.
hartbeeshuis
- Nienaber, G.S.** 1991. Is die modale bywoord *rêrig* *rêrig* Afrikaans? *TGw* 31: 128-40.
rêrig
- Nienaber, G.S.** 1994a. Aantekeninge by *fieterjasies*. *TGw* 34: 64-6.
fieterjasies
- Nienaber, G.S.** 1994b. Kanttekening: *Makrappa*. *TGw* 34: 66.
makrappa
- Nienaber, G.S.** 1995. Die etimologie van *dolos*. *SATT Supp* 23.
dolos; makrappa: 58-9; mamparra: 58
- Odendal, F.F.** 1964. *Karet* en *botel*. *TVV* 20(2): 50-3.
botel: 52-3; karet: 50-2
- Odendal, F.F.** 1966. *Kierang* en *koerang*. *TGw* 6: 151-4.
froetang: 152n; kierang; koerang
- Odendal, F.F.** 1994. Waar die kranse antwoord gee. *FS Raidt*: 29-32.
kranse
- P.** 1911. Het Afrikaanse woord *ou*. *Goede Hoop* November: 393-4.
ou
- Paardekooper, P.C.** 1969. Een Fransvlaams-Zeeuwse parallel: *ons* 'wij'. *TT* 21: 211-5.
ons: 215

- Paardekooper, P.C.** 1990. Een voegwoordraadsel: *Lat 'dat'*. *SATT* 8: 174-7.
lat
- Pauwels, J.L.** 1959a. Afrikaans *hierdie*, *daardie*. *LB Bijblad* 48: 1-3.
daardie; hierdie
- Pauwels, J.L.** 1959b. *FS Pauwels*: 51-70.
oes: 63
- Pienaar, P. de V.** 1928-29. *Kannie-koenie* of *koenie-kannie* (*ghoenie-ghannie*, ens.). *TWK* 10: 99-101.
kannie-koenie
- Rautenbach, W.P.** 1978. Die verlies van die woord *tof* in Afrikaans. *TVV* 34(3): 20-3.
tof
- Roberge, Paul T.** 2001. Diachronic Notes on the Afrikaans Demonstrative Pronouns. *FS vRensburg*: 124-36.
daardie; hierdie
- Roberge, Paul T.** 2002. The Modal Elements *mos* and *maskie* in Cape Dutch. *LS* 24: 397-408.
maskie; mos
- Scholtz, J. du P.** 1963. *Taalhistoriese opstelle. Voorstudies tot 'n geskiedenis van Afrikaans*. Pretoria: J.L. van Schaik.
baie: 210-1(n), 213; *bleik*: 246; *brei*: 182; *daardie*: 138-9; *dalk*: 186; *die*: 134-8; *goël*: 193-4; *hekel*: 234; *hierdie*: 138-9; *hom*: 52-3; *hulle*: 61, 62, 64, 96-101, 105, 111-5; *julle*: 59-64, 82-86; *kaiings*: 171, 209; *karet*: 250; *kasarm*: 240-1; *kê*: 238-9; *koeël*: 234; *lawaa*: 219; *maaier*: 171; *moeg*: 247; *moggel*: 191; *motjie*: 222; *nô*: 211n; *nuut*: 245; *oggend*: 191; *ons*: 93-6, 100-1, 111-2, 115-6; *portfesiede*: 222; *ramkie*: 211n; *rêrig*: 151; *resep*: 219; *rottang*: 211(n); *se*: 52, 119-21; *sens*: 247; *soms*: 219-20; *sooi-brand*: 172; *sproei/spru*: 242; *sulle*: 60-4, 96-8, 113; *tamaai*: 211n; *teiken*: 246; *toiings*: 209; *u*: 64-7, 74-7; *vremd*: 246-7
- Scholtz, J. du P.** 1972. *Afrikaans-Hollands in die agtiende eeu. Verdere voorstudies tot 'n geskiedenis van Afrikaans*. Cape Town: Nasou.
aalwyn: 96; *Afrikaner*: 95; *als*: 102; *askuus*: 91-2; *aspres*: 91-2; *astrak*: 91-2; *astrant*: 94; *baklei*: 82-3; *blaasbalk*: 102; *daardie*: 40; *dalk*: 77; *die*: 35-9; *duwejo*: 50; *els*: 102; *gunter*: 102; *hê*: 55; *het*: 10-1; *hierdie*: 40; *hom*: 79; *kiewiet*: 46; *koejawel*: 98; *likkewa*: 96; *manewales*: 96; *mos*: 79; *olienhout*: 48; *ontwil*: 79; *oond*: 48, 98; *oop*: 78; *pikkewyn*: 96; *prinjeel*: 96; *rasper*: 88; *rinneweer*: 96; *rottang*: 79; *sarlot*: 97; *sens*: 102; *skinder*: 70; *stawel*: 50; *swael*: 48; *swetterjoel*: 96; *tamboekie*: 81; *tameletjie*: 78, 97; *tarentaal*: 67; *tjôwa*: 50; *twak*: 50, 54, 83; *vark*: 78n; *vlei*: 83; *waster*: 94, 95
- Scholtz, J. du P.** 1979. Boerneef se taal. *Burger* 30 May: 12.
bokrin; *bokruit*; *nimlik*
- Scholtz, J. du P.** 1985. Vroeë taalkundige studies. *TGw* 25: 80-105.
ag: 101; *akkerboom*: 94, 100; *burg*: 103; *het*: 97; *karet*: 102; *vremd*: 93
- Schoonees, P.C.** 1950. Aantekening oor die woord *ongans*. *TWK (NR)* 10: 92.
ongans
- Schuchardt, Hugo.** 1885. Review: N. Mansvelt, *Proeve van een Kaapsch-Hollandsch idioticon*. *Literaturblatt für germanische und romanische Philologie* 11: 464-70.
aaapskelder: 466; *arrie*: 468; *balie*: 468; *boeglam*: 466; *katel*: 468-9; *maskie*: 468; *ons*: 468; *poerbale*: 468; *se*: 466; *veels geluk*: 467; *voetsek*: 469
- Schuchardt, Hugo.** 1891. Kreolische Studien IX: Über das Malaioportugiesische von Batavia und Tugu. *Sitzungsgeschichte der Kaiserlichen Akademie der Wissenschaften in Wien, Philosophisch-historische Classe* 122(12): 1-256.
baar: 17, 237; *katel*: 119n; *nô*: 250; *oorkrabbetjie*: 127n; *oorlams*: 16-7; *rabbetjie*: 108n
- Schuhmacher, W.W.** 1973. Once more *ons*. *Standpunte* 28(8): 60.
ons

- Smith, J.J.** 1913. Review: C. Pettman, *Africanderisms*. *Brandwag* 1 December: 394-95, 397-401.
alikeukel: 398; *alla*: 399; *arrie*: 400; *borrie*: 400; *bredie*: 400; *geilsiekte*: 398; *gus*: 398; *harslag*: 399; *kalossies*: 399; *kaparring*: 400; *kapater*: 400; *kardoes*: 400; *kiaat*: 400; *koedoe*: 400; *koejawel*: 400; *kogelman*: 400; *makou*: 400; *manel*: 399; *mielie*: 400; *mombakkies*: 399; *moveer*: 400; *nooi*: 399; *ons*: 397n; *platanna*: 399; *portfesiedeur*: 399; *ramkie*: 400; *rissie*: 400; *sa*: 399; *seekoei*: 399; *skaai*: 399; *swernoot*: 399; *tamaai*: 399; *tasal*: 400; *volstruis*: 400; *weebelaar*: 399
- Smith, J.J.** 1921-22. Notes on G.R. von Wielligh's *Ons geselstaal*. *Huisgenoot* January 1921-July 1922.
 January '21 *bottel*: 393 n. 15; *lemoen*: 393 n. 14; *kapel*: 392 n. 7; *nôï*: 393 n. 12; *perske*: 392 n. 5; *tarentaal*: 393 n. 13; *tiekie*: 393 n. 20; *volstruis*: 392 n. 4; February '21 *pikkenien*: 419 n. 5; March '21 *ghoen*: 457 n. 2; *meestergoed*: 458 n. 8; April '21 *kiets*: 497 n. 5; June '21 *geitjie*: 63 n. 2; July '21 *nôï*: 99 n. 8; *papkuil*: 98 n. 1; *ramkie*: 99 n. 8; *tamaai*: 99 n. 8; August '21 *duiwelsterk*: 146 n. 3; *gramadoelas*: 145 n. 2; *paaboelie*: 145 n. 3; *perske*: 146 n. 1; *seekoei*: 145 n. 1; *voetsek*: 146 n. 8; March '22 *gramadoelas*: 452 n. 1; July '22 *filekommis*: 113 n. 2
- Smith, J.J.** 1938. *Oorlam en baar*. *Stellenbosse Oudstudent* 8(1): 18-23.
baar, *oorlams*
- Smith, J.J.** 1942. 'n Bietjie oor *baie*. *TO*: 173-80.
baie
- Snijman, F.J.** 1977. Nederlandse lexicografiese baanbrekers in Zuid-Afrika. *FS Kruyskamp*: 189-202.
dam: 194; *hakskeen*: 192; *oes*: 193; *vaderlands*: 193-4; *viets*: 192
- Taalman.** 1946. Op taalbesoek by Boerneef. *OEB* 12: 64-68; 143-147.
bokrin: 66; *bokruit*: 66; *skaaprin*: 66; *skaapruit*: 66
- Te Winkel, J.** 1896. Het Nederlandsch in Noord-Amerika en Zuid-Afrika. *VOAT*: 113-134.
alikeukel: 115-116; *amper*: 121; *assegaai*: 124; *baadjie*: 121; *baar*: 121; *baklei*: 122; *baljaar*: 122; *koe-rang*: 123; *kras*: 121; *mielie*: 123; *oorlams*: 121-122; *sambok*: 123-124; *sambreel*: 117; *soebat*: 122
- Valkhoff, Marius F.** 1955. Beskouings oor woordontlening II. *TVV* 12(1-2): 21-9.
bermotsersanpeer: 28n-29n; *dolos*: 23; *kastrol*: 23; *lumier*: 23; *moveer*: 27; *pikkenien*: 27; *tronk*: 27
- Valkhoff, Marius F.** 1969. Review: Boshoff and Nienaber, *Afrikaanse etimologieë*. *AS* 28: 143-6.
baie: 145; *baljaar*: 146; *basta*: 146; *nôï*: 146; *piering*: 146; *pikkenien*: 145; *tamaai*: 146; *tronk*: 146
- Van Blerk, N.** 1958. Oesbeentjie. *Huisgenoot* 28 July: 53.
oesbeentjie
- Van den Berg, B.** 1965. Heeft het Afrikaans zich uit het Hollands ontwikkeld? *FS Scholtz*: 25-38.
die: 32-7
- Van der Meulen, R.** 1917. *Robbedoes*. *TNTL* 36: 1-9.
rabbedoe
- Van Ginneken, J.** 1927. De twee beteekenissen van *kuieren*. *TNTL* 46: 188-9.
kuier
- Van Oordt, J.F.** 1916. Review: D.B. Bosman, *Afrikaans en Maleis-Portugees*. *Brandwag* 1 October: 142-3, 145-9.
amper: 147; *assegaai*: 147; *baie*: 145n; *basaar*: 147; *goiing*: 147; *josie*: 147; *kapater*: 147; *katel*: 147; *koe-doe*: 147; *moveer*: 147; *oorlams*: 147; *piering*: 148; *pondok*: 148; *rissie*: 148; *soebat*: 148
- Van Oordt, L.C.** 1936. Grepe uit die gesproke taal van 1700. *Huisgenoot* 15 May: 27.
booi, *loesing*, *maskie*, *ons*, *sa*, *werda*
- Van Oordt, L.C.** 1941. *Boesman* alias *Bosjesman*. *Huisgenoot* 14 March: 27, 69.
Boesman
- Van Rensburg, Christo.** 1996. Is Koi-invloed op Afrikaans werklik Koi-invloed? *TT Themanummer* 9: 128-45.
ons: 137-41

- Venter, Hertzog.** 1968. Review: Boshoff and Nienaber, *Afrikaanse etimologieë*. *Standpunte* 21(5): 45-7.
ghwarrie: 47; *gus*: 47; *nè*: 47
- Vercoullie, J.** 1919. Negerhollands *molee*, Afrikaans *boetie*, *katjipiering*, *bibies*, *bottel*, *Ou Sanna*, *ewwa-trewwa*, *foolstruis*. *TNTL* 38: 302-6.
biebies: 305; *boet*: 304; *bottel*: 305; *ewwa-trewwa*: 305; *katjipiering*: 304; *ousanna*: 305; *vieruurtjie*: 305; *volstruis*: 306
- Von Wielligh, G.R.** 1921-22. Ons geselstaal. *Huisgenoot* January 1921–July 1922.
March '21 *pasella*: 457; July '21 *baie*: 98-9; *papkuil*: 98; August '21 *agostou*: 145-6; *beesbyter*: 145; *eendjies*: 145; *gansies*: 145; *kankerbossie*: 145; *makoppa*: 145; *perske*: 146; *seekoei*: 145; *watter*: 146; September '21. *herneuters*: 200
- Vor der Hake, J.A.** 1911. Is de beleefdheidsvorm *U* 'n verbastering van *UEd*? *NT* 5: 16-24.
hulle: 19; *ons*: 19; *u*
- Waher, Hester.** 2001. Woorde wat wegraak: oorgeërfde Franse leenwoorde in Afrikaans. *FS vRensburg*: 157-63.
menteneer: 157-8; *resnawel*: 158-9; *spandabel*: 161-2
- Walton, James.** 1987. *Hartbeeshuis* and *hartbeeshut*. *TVV* 43(1): 15-23.
hartbeeshuis

Lys Afkortings / List of Abbreviations

- Acta Germanica* = *Acta Germanica. Zur Sprache und Dichtung Deutschlands, Österreichs und der Schweiz* (Cape Town)
- AN&N* = *Africana Notes and News* (Johannesburg)
- AS* = *African Studies* (Johannesburg)
- Brandwag* = *Die Brandwag* (Pretoria)
- Burger* = *Die Burger* (Cape Town)
- FS HG* = *Festschrift dem Hansischen Geschichtsverein und dem Verein für niederdeutsche Sprachforschung dargebracht zu ihrer Jahresversammlung in Göttingen Pfingsten 1900*. Göttingen: Dieterichsche Univ-Buchdruckerei von W. Fr. Kästner.
- FS Louw* = *Smal swaard en blink. Bundel aangebied aan N.P. van Wyk Louw by geleentheid van sy sestigste verjaardag, 11 Junie 1966*. Pretoria/Cape Town: Academica.
- FS Kempen* = *Taalkunde — 'n Lewe. Studies opgedra aan Prof. Dr. W. Kempen by geleentheid van sy 65ste verjaardag, 3 Julie 1974*. Odendal, F.F., D.C. Hauptfleisch, J.E. Loubser (Eds.). Cape Town/Johannesburg: Tafelberg.
- FS Kruyskamp* = *Opstellen door vrienden en vakgenoten aangeboden aan Dr. C.H.A. Kruyskamp ter geleentheid van zijn 65ste verjaardag en van zijn afscheid als redacteur van het Woordenboek der Nederlandsche Taal en als enige bewerker van Van Dale*. Heestermans, Hans (Ed.). The Hague: Martinus Nijhoff.
- FS Nienaber* = G.S. Nienaber — 'n huldeblyk. *Studies opgedra aan Prof. Dr. G.S. Nienaber in sy tagtigste jaar*. Sinclair, A.J.L. (Ed.). 1983. Bellville: University of the Western Cape.
- FS Pauwels* = *Verzamelde Opstellen. Aangeboden aan de Auteur ter Gelegenheid van zijn 65e Verjaardag*. Engels, L.K., J. van Haver, G.Geerts (Eds.). 1965. Assen: Van Greunen.
- FS Raidt* = *Nuwe perspektiewe op die geskiedenis van Afrikaans: opgedra aan Edith H. Raidt*. Olivier, Gerit, Anna Coetzee (Eds.). 1994. Johannesburg: Southern Book Publishers.
- FS Scholtz* = *Dietse studies. Bundel aangebied aan Prof. Dr. J. du P. Scholtz by geleentheid van sy vyf-en-sestigste verjaardag 14 Mei 1965*. Lindenberg, E., Edith Raidt, J.A. Verhage (Eds.). Cape Town/Pretoria: Academica.

- FS vdMerwe* = *Gedenkbundel H.J.J.M. van der Merwe*. De Klerk, N.J., F.A. Ponielis (Eds.). 1976. Pretoria: J.L. van Schaik.
- FS vRensburg* = *Taallandskap. Huldigingsbundel vir Christo van Rensburg*. Carstens, Adelia, Heinrich Grebe (Eds.). 2001. Pretoria: Van Schaik.
- FS vSchalkwyk* = *'n Man wat beur. Huldigingsbundel vir Dirk van Schalkwyk*. Botha, Willem (Ed.). 2003. Stellenbosch: Bureau of the WAT.
- Goede Hoop* = *De Goede Hoop* (Cape Town)
- Handhaaf* = *Handhaaf* (Johannesburg)
- HS* = *Historiese Studies* (Pretoria)
- Huisgenoot* = *Die Huisgenoot* (Cape Town)
- IJGLSA* = *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* (Berkeley)
- Koers* = *Koers. Tweemaandelikse tydskrif* (Potchefstroom)
- LB* = *Leuvense Bijdragen* (Louvain)
- LS* = *Language Sciences* (Oxford)
- NT* = *De Nieuwe Taalgids* (Groningen)
- OEB* = *Ons Eie Boek* (Cape Town)
- Raidt 1994* = *Raidt, Edith H. 1994. Historiese taalkunde. Studies oor die geskiedenis van Afrikaans*. Johannesburg: Witwatersrand University Press.
- SATT* = *Suid-Afrikaanse Tydskrif vir Taalkunde/South African Journal of Linguistics* (Pretoria)
- SATT Supp* = *Suid-Afrikaanse Tydskrif vir Taalkunde/South African Journal of Linguistics. Supplement/Supplement* (Pretoria)
- Standpunte* = *Standpunte* (Cape Town)
- Taalfasette* = *Taalfasette* (Pretoria)
- TGw* = *Tydskrif vir Geesteswetenskappe* (Pretoria)
- T&L* = *Taal en Letteren* (Zwolle)
- TNTL* = *Tijdschrift voor Nederlandse Taal- en Letterkunde* (Leyden)
- TO* = *Taalkundige opstelle*. P.J. Nienaber (Ed.). 1965. Cape Town/Amsterdam: A.A. Balkema.
- TT* = *Taal en Tongval* (Amsterdam/Ghent)
- TVV* = *Tydskrif vir Volkskunde en Volkstaal* (Johannesburg)
- TWK* = *Tydskrif vir Wetenskap en Kuns* (Pretoria)
- TWK (NR)* = *Tydskrif vir Wetenskap en Kuns (Nuwe Reeks)* (Pretoria)
- Vaderland* = *Die Vaderland* (Johannesburg)
- VOAT* = *Vroeë opvattinge oor Afrikaans as taal*. G.S. Nienaber (Ed.). 1984. Johannesburg/Cape Town: Perskor.
- Volkstem* = *De Volkstem* (Pretoria)

Un vent nouveau pour les éditions Robert

Jean-Nicolas De Surmont, *University of Leeds, Leeds, United Kingdom*
(jdesurmont@yahoo.fr)

Après la réédition de ses produits électroniques (*Petit Robert électronique* 2007 paru en septembre 2006 et le *Grand Robert électronique* paru en 2005) mentionnons la récente parution de deux ouvrages: *Le Dictionnaire des combinaisons de mots: les synonymes en contexte* (collection les Usuels, 2007, 1011 pages) et le *Robert Van Dale Dictionnaire français-néerlandais; néerlandais-français*. Il convient néanmoins de signaler que l'une des particularités des récentes éditions Robert consistera en l'ajout de 48 belgicisms s'appuyant en partie sur un travail que nous avons effectué au Centre de recherche Valibel en 2003 et 2004.[†] Cette nouveauté belge dans le *Petit Robert* fait suite à notre suggestion aux éditions Robert en 2003 de revoir la nomenclature des belgicisms dans les dictionnaires Robert. En outre, à la demande de Madame Durand, récemment nommée à la direction des éditions Robert, et qui voulait se mettre au goût du jour en vue du festival de la Francophonie de 2006 en France, le Centre de recherche Valibel a été prié de livrer le fruit de ses recherches entamées depuis 2003 et qui a fait l'objet de rencontres nombreuses. Le Festival de la Francophonie constituait en quelque sorte une réaction contre des commentaires selon lesquels la France s'excluerait du mouvement de la Francophonie. (Voir notre article paru sur <http://www.lianes.org>). Le *Petit Larousse illustré* a tenté de renverser la tendance en incluant la France dans les activités de la Francophonie en 2006.

Cette volonté de prendre en compte les topolectismes est également présente dans le *Dictionnaire bilingue français-néerlandais/néerlandais-français* dirigé par Paul Bogaards et qui en est à sa quatrième édition. La publication d'un dictionnaire bilingue de ce type nous paraît primordiale dans un pays comme la Belgique, dont les deux langues les plus importantes administrativement parlées sont le français et le flamand, ainsi qu'en Hollande où la pénétration du français remonte, du moins dans les classes dominantes à plusieurs siècles, et où sont publiés au XVII^e siècle, les dictionnaires de Furetière et Bayle. Parmi les 700 mots nouveaux de ce dictionnaire, dont la première édition remonte à 1988, mentionnons l'ajout de flandricisms, qu'on appelle parfois à tort *belgicisme* alors que dans la terminologie métalinguistique il s'agit en fait de particularités francophones. On note dans le flamand de Belgique quelques flandricisms comme les toponymes (*Rijsel* pour *Lille* par exemple). C'est surtout le néerlandais qui a été mis à jour compte tenu de la nouvelle orthographe, quant à la partie française, notamment pour les belgicisms, on a considéré que le fonds du *Petit Robert* était suffisant.

Contrairement au dictionnaire *Robert Collins*, le dictionnaire bilingue français/néerlandais est monodirectionnel et s'adresse essentiellement aux francophones de Belgique. *Le Dictionnaire Van Dale/Robert*, très connu, s'est vendu à 10 000 exemplaires pour sa première édition, essentiellement en Belgique, alors que le *Petit Robert* se vend annuellement à environ 200 000 exemplaires en France. Parmi les ajouts en langue française mentionnons *courriel*, *arobase*, *chat*, *texto*, *teleachat*, *e-commerce*, *placoplâtre* et du côté flamand: *drukkingsgroep*, *schoollopen*, *inpikken op*, *onthaal*, *bissen*, traduit par la métalangue belge ce qui est exceptionnel. L'ajout de *courriel* nous semble intéressant puisqu'il témoigne de la reconnaissance d'une recommandation québécoise à l'échelle internationale au détriment de *Mél* qui est en perte d'usage.

Récemment présenté au Salon du livre de Paris de 2007, *Le Dictionnaire de combinaisons de mots* des éditions Robert est paru le 15 mars 2007. Il est le premier dictionnaire de ce type depuis 250 ans selon l'affirmation de Dominique Le Fur qui a dirigé l'ouvrage. Il comporte 160 000 combinaisons de mots classées par sens accompagnées de 26 000 exemples. Il s'agit en fait de syntagmes à figement assez important dans la langue comme *santé de fer*, *faire entorse à une loi*, *candidat de poids*, *traduire en justice*, *harcèlement policier*, *discrimination institutionnelle*, etc. Le dictionnaire a nécessité trois ans de travail et s'appuie sur un corpus de 500 000 millions de mots-occurrences.

Parmi les prochaines productions des éditions Robert figurera l'édition de poche du bilingue néerlandais/français et la remise à jour du Super Senior Collins/Robert.

Mentionnons que la Belgique a non seulement fait un accueil favorable aux récents dictionnaires Robert mais également à la publication du *Dictionnaire des Belgicisms* de Gorges Lebouc (Bruxelles, Éditions Racine, 2006, préface d'Henriette Walter), et qui s'appuie sur un corpus tiré du journal *Le Soir* et présentant parfois les emplois similaires de certains belgicisms en France et au Québec. Dans ce cas de nombreuses réserves sont à émettre. La présentation des belgicisms en contexte, dans un texte à thème bien ciblé, nous semble fort intéressant sur le plan pédagogique.

Note

- † Le site du Centre de recherche Valibel à Louvain-la-Neuve, dirigé par Michel Francard, indique que cela est paru dans l'édition 2007 du *Petit Robert* mais c'est en fait de l'édition 2008 qui paraîtra en août 2007 qu'il s'agit. En effet en cette année 2006 ne sont ajoutés à la nomenclature du *Petit Robert* que quelques belgicisms.

B.T. Sue Atkins Receives an Honorary Doctorate from the University of Pretoria

On 5 September 2008, the degree of Doctor of Literature, *Honoris Causa*, was conferred on B.T. Sue Atkins by the University of Pretoria. The decision by the Council and the Senate to award her this degree was taken on the grounds of the following considerations as stated in the following commendatio read at the graduation ceremony.

Sue Atkins has been a professional lexicographer since 1966, first with Collins (now HarperCollins) Publishers where she was General Editor of the first 'modern' innovative English–French dictionary, the *Collins-Robert English–French Dictionary*, then as Lexicographic Adviser to Oxford University Press, where she pioneered an effective methodology for the creation of bilingual dictionaries from corpus data, resulting in the *Oxford-Hachette English–French Dictionary*. She is currently Lexicographic Adviser to the FrameNet project at the International Computer Science Institute, Berkeley, California, and a member of the Advisory Board of the American National Corpus, and the *International Journal of Lexicography*.

In 2000, she was awarded an honorary D.Litt. by the University of Brighton, United Kingdom, for services to lexicography and linguistics, national and international. In 2002, she received a festschrift published by the European Association for Lexicography (EURALEX) to mark her contribution to international lexicography, and was made an Honorary Life Member of EURALEX, of which she is a past President.

She has organized and taught at many professional and academic training courses and workshops in lexicography, and held lexicographic consultancies with a number of prestigious research institutions. She has participated in important national and international research projects in the field of computational lexicography, and originated the idea of the British National Corpus.

Her main focus includes the lexical analysis of corpus data, and in particular the use of linguistic theory as basis for a systematic description of the language; designing databases to store lexicographic data for use by human lexicographers and computer lexicons; applying such databases in the creation of monolingual and bilingual dictionaries; the training of lexicographers; and the study of how dictionaries are actually used.

Sue Atkins presented more than 40 papers and presentations and is author or co-author of numerous publications of which *The Oxford Guide to Practical Lexicography* (2008) in cooperation with Michael Rundell is the latest.

Her influence in the lexicographic world is described as follows in the foreword to *Lexicography and Natural Language Processing: A Festschrift in Honor of B.T.S. Atkins* (2002): 'Sue Atkins has not only inspired generations of lexicographers, but has also played a crucial role in developing links between the practical world of dictionary-making and the research communities of linguistics and natural language processing.'

With her workshops SALEX '97 (hosted by Rhodes University, Grahamstown) and AFRILEX-SALEX '98 (hosted by the University of Pretoria, Pretoria), Sue Atkins made an invaluable contribution to the development of lexicography in South Africa, in particular for Sepedi, Setswana, Sesotho, isiNdebele, isiZulu, Siswati, isiXhosa, Tshivenda and Xitsonga. As a result of these groundbreaking events and the subsequent widespread interest in designing and producing dictionaries for these languages, dictionary-making in South Africa has shifted from random, haphazard publications to a serious commitment to lexicographic principles and practice. More and more dictionaries of a high quality are now being produced in South Africa. Sue Atkins receives an honorary doctorate from the University of Pretoria for her outstanding achievements and contributions to international lexicography and lexicographic guidance to South Africans.

At the graduation ceremony, Atkins responded with the following speech of acknowledgement:

It is with the greatest pleasure, and a sense of deep humility, that I accept the honour you confer on me today, thus belying the words of our great predecessor in lexicography, Samuel Johnson, who said in the *Preface* to his splendid *Dictionary* of 1755:

It is the fate of those who toil at the lower employments of life ... to be exposed to censure, without hope of praise; to be punished for neglect, where success would have been without applause, and diligence without reward. Among these unhappy mortals is the writer of dictionaries ... Every other authour may aspire to praise; the lexicographer can only hope to escape reproach ...

I have believed this for over half a lifetime, and I am now in a state of shock. And yet I realise that this honour marks, not so much my own contribution to dictionary making, as the great surge in lexicographic activity and enthusiasm which has swept over South Africa during the past ten years. Your government is to be applauded for its far-sightedness in legislating for the nationwide dictionary-making initiative, and thus opening the way to all this exciting work that is happening around us here. It was, I know, thanks to these measures that your National Lexicographic Units (NLUs) were established, and I know, too, that one of the catalysts in this whole movement was the Department of African Languages in the University of Pretoria. It is an honour and a pleasure to be associated with this great institution.

I first came to South Africa in 1997, to lead the SALEX (South African Lexicography) workshop, held at Rhodes University, Grahamstown, under the auspices of the Dictionary Unit for South African English, and organized by its Director. This first workshop lasted two weeks, and was attended by almost forty South African linguists, many of them distinguished scholars, and some already with considerable lexicographic experience. During these two weeks, we worked through the basics of dictionary editing, and the various language groups each produced one or two dictionary entries. These were entries for a monolingual dictionary. Bilingual dictionaries present different problems and require different training, as was realised by the linguists who founded AFRILEX, the African Association for Lexicography. And so the following year I returned to South Africa, this time to Pretoria, with my colleague Michael Rundell, to lead the AFRILEX-SALEX Workshop in bilingual dictionary-making, organized by Professor D.J. Prinsloo, of the Department of African Languages. This time, over fifty South African linguists spent two weeks learning the trade and writing entries from English into their own language. The most striking thing about these events was the extraordinary energy of the participants, and their unquenchable enthusiasm for everything lexicographic. The two Workshops, I believe, provided the basic training of the South African linguists who were subsequently to work on the various African language dictionaries in the newly-founded Dictionary Units, as well as those working in other lexicographic initiatives. I know that the University of Pretoria is proud to house the isiNdebele Unit, and a branch office of the Sepedi Unit.

Now I would like to mention the difference between dictionary writing as it used to be, and dictionary writing today. Until recently, lexicographers collected citations on index cards or in notebooks, and wrote dictionary entries on the basis of these notes together with what they knew about the word. With the advent of computers, however, came the *text corpus*. This is an electronic collection of texts, drawn from books, letters, newspapers, transcribed conversation and so on. These texts are selected and prepared for use by lexicographers, with software that allows us to look at thousands of instances of a word in real use, that is, as it is used by ordinary people in everyday life. Such a corpus takes us beyond what we as individuals know about the word, and we no longer have to rely solely on how we think the word is used. This is an invaluable tool for dictionary editors, and sometimes it is just plain fascinating as well. For instance, when I ask English speakers to give me an example of the use of the word *ago*, they will say 'fifty years ago' or 'a couple of days ago' or 'a long time ago'. Twenty-five years ago, when I first looked this word up in a corpus, I was delighted to find phrases like 'three ex-wives ago' and 'four gins and two martinis ago'. We would never have thought them up, yet we can all understand them. Seriously, however, nowadays most respectable dictionary projects start from an analysis of corpus data.

The first initiative in corpus building in South Africa came from the University of Pretoria. Already in 1989, the Department of African Languages had

seen the start of corpus creation for the African languages spoken in South Africa, that is to say, Sepedi, Setswana, Sesotho, isiNdebele, isiZulu, Siswati, isiXhosa, Tshivenda and Xitsonga. This work received a great impetus with the advent of the National Lexicographic Units, who took the principled decision to consult text corpora for evidence on how the words of their language behave. Since then these corpora have also become the basis of lexicographic and linguistic research, the development of commercial spell-checkers, of part-of-speech taggers, of lexicons for machine translation, and other similar applications. These South African corpora are mainly texts scanned from literary works, both prose and poetry, and also grammar books, newspapers and magazines. A start has been made on compiling 'oral corpora', which contain transcriptions of spontaneous and semi-scripted speech. These nine corpora range in size from 2 million to 11 million words and they continue to grow. This is a tremendous achievement for a single nation.

To set this achievement in context, the first British lexicographic corpus freely available to all researchers, the British National Corpus, took three years to complete, involved the collaboration of three major dictionary publishers, two universities and the British Library, and 15 years ago cost one million pounds, half of which came from the government. The American National Corpus, since its inception in 1999, has so far managed to collect and process only 22 million words of American English, despite its enormous geographical coverage, the many universities and publishers involved, and the very many works now available in electronic format, thanks to computer typesetting. When set against these facts, the South African corpus-building initiative is truly impressive.

The long-term goal of the legislation was the production of a monolingual dictionary in each of the nine African languages, and this objective lies at the heart of the academic work being carried out by Lexicographic Units all over the Republic of South Africa. Quite a number of basic monolingual dictionaries have been published, and some Units have also published bilingual dictionaries. The international Laurence Urdang Award to support lexicographic work, won by Professor Prinsloo and the Department of African Languages, was used to bring together the consultative body to establish the lexicographic needs of isiNdebele.

Other initiatives have arisen directly or indirectly out of the AFRILEX-SALEX Workshop. Scholarly writings based on analyses of corpus data have established South Africa as a focus of lexicographic work in the eyes of the international research community, and this work continues to grow and involve a new generation of South African linguists. Bilingual dictionaries from English into local African languages, have begun to appear as commercial publications. An excellent software suite for the production of dictionaries, TshwaneLex, created and first trialled in South Africa, is now in use throughout the world.

Despite this encouraging situation, many challenges still lie ahead. Unlike the major institutions in Europe and the U.S.A., with a long tradition of scholarly and historical lexicographic work on linguistic principles, the Dictionary Units for the African languages of South Africa have to build their own lexicographic tradition, and this is not something that can happen overnight, although the recent work by the members of AFRILEX has laid down a solid foundation on which to build your modern dictionaries. For this great lexicographic initiative to continue to grow, develop and prosper, there is a need for a united effort on the part of your institutions, including the University of Pretoria, to bring dictionaries to the forefront of linguists' concerns, so that not only university academics and researchers, but school teachers and students of all ages, may appreciate the new dictionaries and use them to the full. And, of course, continue to record and refine the account of the languages to be found in them.

To that end, I would urge you to focus on teaching dictionary skills to students at schools and universities (and perhaps to their teachers too); to encourage scientifically planned research into the use of dictionaries in schools, colleges, universities and the home; and to make your dictionaries as user-friendly as possible. Here I should stress my belief that if a careful student makes a mistake when using his/her dictionary, then in most cases that is the fault of the dictionary not the student. A dictionary is a dialogue between the lexicographer and the dictionary user. We can write clear, accurate and helpful entries that suit our target readership without lowering lexicographic standards. This practice is already firmly established in South Africa: long may it prosper!

D.C. Hauptfleisch 14 Maart 1926 – 14 Junie 2008

Op 1 Junie 1958 het D.C. Hauptfleisch as opsteller (later mederedakteur) lid geword van die redaksie van die *Woordeboek van die Afrikaanse Taal* (WAT) onder dr. P.C. Schoonees, die tweede hoofredakteur. Op 1 Oktober 1966 is hy bevorder tot assistenthoofredakteur. Na die aftrede van die derde hoofredakteur, dr. F.J. Snijman, aan die einde van 1980, is hy met ingang 1 Mei 1981 as vierde hoofredakteur van die WAT aangestel. Hy het hierdie pos beklee tot sy aftrede op 31 Julie 1991 toe *Lexikos* as huldigingsbundel aan hom oorhandig is. Hy is op 14 Junie vanjaar in die ouderdom van 82 jaar oorlede.

Namens huidige en oudkollegas word hiermee deelneming aan sy familie betuig: aan sy vrou Maria, sy vier kinders Linda, Stefan, Therese en Sonja, asook sy ses kleinkinders. En terselfdertyd word dankie gesê aan hulle as familie wat Niël met die WAT wat so 'n integrerende deel van sy lewe gevorm het, moes deel. Hy het dikwels tot laataand, oor naweke en selfs in sy vakansietye gewerk. Gedurende sy totale dienstyl van meer as 33 jaar, was die WAT 'n belangrike deel van sy bestaan.

Daniël Carinus Hauptfleisch is op 14 Maart 1926 in die Paarl gebore waar hy op die plaas Lemoenkloof grootgeword het. Vanaf 1935 het hy op Stellenbosch skoolgegaan waar hy in Desember 1943 die Matrikulasie-eksamen aan die Hoër Seunskool Stellenbosch (tans Paul Roos Gimnasium) in die eerste klas met onderskeiding in Afrikaans geslaag het. Sy Latynonderwyser was die legendariese Paul Roos self. Aan die einde van 1946 behaal hy die B.A.-graad cum laude aan die Universiteit van Stellenbosch met Afrikaans-Nederlands en Latyn as hoofvakke. Gedurende 1948 en 1949 studeer hy verder aan die Universiteit van Stellenbosch vir die M.A.-graad in Afrikaans-Nederlands. In 1953 verwerf hy hierdie graad cum laude met 'n skripsie *Vorm en gebruik van attributief-adjektiewiese voltooidedeelwoorde in Afrikaans, met inbegrip van skynbare partisipia* wat hy onder leiding van prof. dr. W. Kempen voltooi het. Behalwe vir die tydperk 1 Januarie 1950 tot 31 Mei 1958 toe hy as vertaler in diens van die destydse Volksraad in Kaapstad was, het Hauptfleisch die grootste deel van sy lewe op Stellenbosch deurgebring.

Gedurende September 1972, toe assistenthoofredakteur, besoek hy die Instituut voor Nederlandse Lexicologie (INL) (Thesaurus en WNT) in Leiden en die OED Supplement in Oxford waar hy verskillende aspekte en probleme betreffende die redaksionele werkwyse, materiaalversameling en leksikografiese apparaat met lede van die onderskeie redaksies bespreek het. Na sy aanstelling as hoofredakteur het hy hierdie kontak voortgesit. Gedurende sy redakteurskap het die Buro institusionele lidmaatskap van die European Association for Lexicography (EURALEX) en van die Dictionary Society of North

America (DSNA) verkry wat meer uitgebreide leksikografiese kontak moontlik gemaak het. Voortdurende skakeling is ook volgehou met die Rhodes University English Dictionary Unit in Grahamstad waar die *Dictionary of South African English on Historical Principles* saamgestel is, destyds, behalwe die Buro van die WAT, die enigste ander leksikografie-eenheid in Suid-Afrika.

Nadat hy hoofredakteur geword het, het Hauptfleisch homself daarvoor beywer om die werk aan die WAT te versnel deur die leksikografiese proses meer vaartbelyn te maak. Hiervan getuig sy referaat "Die Woordeboek van die Afrikaanse Taal: Verlede, hede en toekoms", wat hy in 1982 gedurende sy vroeë jare as hoofredakteur op die Agtiende Nasionale Kongres van die Linguistevereniging van Suider-Afrika by die Rhodes Universiteit, Grahamstad, gelewer het. In hierdie aanbieding konsentreer hy veral op die redaksionele beleid deur eers 'n oorsig van die verlede en hede te gee voordat hy aandui wat vir die toekoms beoog word om die redaksionele proses ten opsigte van die produksietempo en die alfabetiese vordering te versnel. Hy beklemtoon die belangrikheid van voortdurende kontak met die gebruikers van die WAT om hul menings en behoeftes te bepaal.

Met groot noukeurigheid het hy die redigering van die manuskrip gehanteer, die redaksionele stelsel op datum gehou en die bronverwysings versorg. Onder sy hoofredakteurskap het deel VII van die WAT verskyn. Deel VIII wat volledig deur hom geredigeer is, is kort na sy aftrede op 16 November 1991 gepubliseer.

Hauptfleisch was toeganklik vir gemotiveerde voorstelle vir die verbetering en versnelling van die werksaamhede van die Buro. Teen die einde van sy dienstermyn is met strategiese beplanning begin wat gelei het tot 'n nuwe redaksionele beleid waarvan die uitvoering met behulp van 'n paslike rekenaarsstelsel 'n belangrike verandering en 'n aanmerklike versnelling van die werk aan die WAT teweeggebring het.

In 1968 het hy lid van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns geword en in 1971 lid van die Taalkommissie van die Akademie wat belas is met die samestelling van die *Afrikaanse Woordelys en Spelreëls*. Later, as sekretaris, was hy verantwoordelik vir die noukeurige en tydrowende notulering van die besluite geneem tydens die weeklange sittings van die Taalkommissie.

Hauptfleisch se taalkundige en leksikografiese belangstellings het verskeie publikasies tot gevolg gehad. Tussen 1967 en 1969 het hy, deur gebruikmaking van materiaal wat hy vir 'n beplande proefskrif versamel het, drie artikels in onder meer die tydskrif *Standpunte* oor woordherhalings- en reduplikasiever-skynsels in Afrikaans gepubliseer. In 1974 was hy saam met F.F. Odendal en J.E. Loubser mederedakteur van 'n bundel studies onder die titel *Taalkunde — 'n lewe*, opgedra aan prof. dr. W. Kempen by geleentheid van sy 65ste verjaarsdag. In *Spil Plus* 10 (1984), die tydskrif van die Departement Algemene Taalwetenskap aan die Universiteit van Stellenbosch, het 'n artikel oor "Taalnormering en woordeboeke: 'n Praktykgerigte perspektief" verskyn. In April 1989 het

hy deelgeneem aan 'n leksikografiesimposium by die Departement Afrikaans en Nederlands aan die Universiteit van Stellenbosch waar hy die onderwerp "The Definition of Meaning in Overall Descriptive Dictionaries" ingelei het. Maar dit was veral vanweë die veranderende sosiaal-politieke situasie in Suid-Afrika dat hy in artikels aandag begin skenk het aan beledigende en sensitiewe taalgebruik, byvoorbeeld in die behandeling van rassistiese en seksistiese terme in woordeboeke. 'n Artikel oor "Taalseksisme en die woordeboek" is opgeneem in die huldigingsbundel geredigeer deur T.J.R. Botha, *Leksikografie en leksikologie* wat aan prof. dr. F.F. Odendal in September 1989 by sy aftrede aangebied is. Die artikel "Racist Language in Society and in Dictionaries: A Pragmatic Perspective" wat in *Lexikos* 3 (1993) verskyn het, is in dié verband 'n belangrike bydrae wat die basis gevorm het vir die WAT se beleid oor die definiëring van sulke terme.

Na sy aftrede het hy in samewerking met P. Hartevelt en L.G. de Stadler die *Woordkeusegids — 'n Kerntesourus van Afrikaans* saamgestel wat in 1992 verskyn het. Hy het ook gedurende die jare 1994–1996 as resensieredakteur van *Lexikos* 4–6 opgetree. In 1995 het hy 'n erepenning vir leksikografiese werk van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns ontvang.

As hoofredakteur het kollegas hom leer ken as 'n persoon met onvermoeide werkkrag en absolute presiesheid, gepaardgaande met 'n sin vir billikheid, toeganklikheid en beleefdheid. Na sy aftrede het hy sy volgehoue belangstelling in en meeewing met die werksaamhede van die Buro en die Taalkommissie getoon. Kollegas kan daarvan getuig dat Niël soos in Jiddisj gesê word, 'n "mensch" was: 'n onbaatsugtige persoon gekenmerk deur eerbaarheid, opregtheid en meelewendheid. Hy het Afrikaans só gepraat en geskryf dat dit sy geldigheid as medium in alle taalsituasies gedemonstreer het. Vir daardie toewyding en vir die mens wat hy was, word sy gedagtenis geëer.

D.C. Hauptfleisch 14 March 1926 – 14 June 2008

On 1 June 1958, D.C. Hauptfleisch joined the editorial staff of the *Woordeboek van die Afrikaanse Taal* (WAT) as compiler (later co-editor) under Dr P.C. Schoonees, the second editor-in-chief. On 1 October 1966, he was promoted to assistant editor-in-chief. After the retirement of the third editor-in-chief, Dr F.J. Snijman at the end of 1980, he was appointed fourth editor-in-chief of the WAT as from 1 May 1981. He held this position until his retirement on 31 July 1991 when *Lexikos* was presented to him as festschrift. On 14 June this year, he died at the age of 82 years.

On behalf of present and former colleagues, sympathy is herewith extended to his family: to his wife Maria, his four children Linda, Stefan, Therese and Sonja, as well as his six grandchildren. And at the same time thanks is given to the family who had to share Niël with the WAT which formed such a part and parcel of his life. He often worked until late at night, at weekends and even during his holidays. During his total tenure of office of more than 33 years, the WAT was an important part of his life.

Daniël Carinus Hauptfleisch was born on 14 March 1926 in the Paarl where he grew up on the farm Lemoenkloof. From 1935, he attended the Boys' High School (now Paul Roos Gymnasium) in Stellenbosch where he passed the Matriculation Examination in the first class with distinction in Afrikaans in December 1943. His Latin teacher was the legendary Paul Roos himself. At the end of 1946, he obtained the B.A. degree cum laude at the University of Stellenbosch, majoring in Afrikaans-Dutch and Latin. During 1948 and 1949, he continued his studies at the University of Stellenbosch for the M.A. degree in Afrikaans-Dutch. In 1953, he obtained this degree cum laude, his thesis, completed under the guidance of Prof. Dr W. Kempen, dealing with *Vorm en gebruik van die attributief-adjektiewiese voltooid deelwoorde in Afrikaans, met inbegrip van skynbare partisipia (Form and Use of the Attributive Adjectival Past Participles in Afrikaans, Including Apparent Participia)*. Except for the period 1 January 1950 to 31 May 1958 when he was in the service of the then House of Assembly in Cape Town as a translator, Hauptfleisch spent the greatest part of his life in Stellenbosch.

During September 1972, at that time assistant editor-in-chief, he visited the Instituut voor Nederlandse Lexicologie (INL) (Thesaurus and WNT) in Leyden and the OED Supplement in Oxford where he discussed various aspects and problems regarding the editorial modus operandi, the collection of material and the lexicographical apparatus with members of the different editorial staffs. After his appointment as editor-in-chief, he continued this contact. During his editorship, the Bureau acquired constitutional membership of the Euro-

pean Association for Lexicography (EURALEX) and the Dictionary Society of North America (DSNA), which made more extensive lexicographical contact possible. Continuous liaison was also maintained with the Rhodes University English Dictionary Unit in Grahamstown, where the *Dictionary of South African English on Historical Principles* was compiled, at that time, except for the Bureau of the WAT, the only other lexicography unit in South Africa.

After his appointment as editor-in-chief, Hauptfleisch aimed at speeding up the work on the WAT by making the editorial process more streamlined. Bearing witness to this, is his paper 'Die Woordeboek van die Afrikaanse Taal: Verlede, hede en toekoms' which he read in 1982, during his early years as editor-in-chief, at the Eighteenth National Congress of the Linguistic Society of Southern Africa at Rhodes University, Grahamstown. In this presentation, he particularly concentrates on editorial policy by first giving an overview of the past and the present before indicating what is envisaged for the future to accelerate the editorial process with regard to production rate and alphabetical progress. He accentuates the importance of continuous contact with the users of the WAT to determine their views and needs.

With great accuracy, he handled the editing of the manuscript, kept the editorial system up-to-date and attended to the source references. During his tenure as editor-in-chief, volume VII was published. Volume VIII, which he fully edited, appeared on 16 November 1991 shortly after his retirement.

Hauptfleisch was accessible to motivated suggestions for the improvement and acceleration of the activities of the Bureau. Towards the end of his office, strategic planning was started, which resulted in a new editorial policy which, with the implementation of a suitable computer system, brought about an important change in and a considerable acceleration of the work on the WAT.

In 1968, he became a member of the South African Academy for Science and Arts, and in 1971, a member of the Language Committee of the Academy, that is charged with the compilation of the *Afrikaanse Woordelys en Spelreëls*. Later, as secretary, he was responsible for the accurate and time-consuming recording of the decisions taken during the week-long sessions of the Language Committee.

Hauptfleisch's linguistic and lexicographical interests resulted in several publications. Between 1967 and 1969, using material he collected for a planned doctoral thesis, he published three articles in, among others, the journal *Standpunte*, dealing with word repetition and reduplication features in Afrikaans. In 1974, he was the co-editor together with F.F. Odendal and J.E. Loubser of a collection of studies with the title *Taalkunde — 'n lewe*, dedicated to Prof. Dr W. Kempen on the occasion of his 65th birthday. In *Spil Plus* 10 (1984), the journal of the Department of General Linguistics at the University of Stellenbosch, an article appeared about 'Taalnormering en woordeboeke: 'n Praktykgerigte perspektief' (Language Standardization and Dictionaries: A Practice-directed Perspective'). In April 1989, he took part in a lexicography symposium at the

Department of Afrikaans and Dutch at the University of Stellenbosch where he introduced the subject 'The Definition of Meaning in Overall Descriptive Dictionaries'. But it was especially because of the changing social-political situation in South Africa that he started directing his attention in articles towards offensive and sensitive language use, for example in the treatment of racist and sexist terms in dictionaries. An article on 'Taalseksisme en die woordeboek' ('Language Sexism and the Dictionary') is included in the festschrift edited by T.J.R. Botha, *Leksikografie en leksikologie*, which was presented to Prof. Dr F.F. Odendal on his retirement in September 1989. The article 'Racist Language in Society and in Dictionaries: A Pragmatic Perspective' that appeared in *Lexikos* 3 (1993) is in this regard an important contribution, which formed the basis for the policy of the WAT on the defining of such terms.

After his retirement, he compiled, in co-operation with P. Harteveld and L.G. de Stadler, the *Woordkeusegids — 'n Kerntesourus van Afrikaans* that appeared in 1992. Also during the years 1994–1996, he acted as review editor of *Lexikos* 4–6. In 1995, he received a medal of honour for lexicographical work from the South African Academy for Science and Arts.

As editor-in-chief, colleagues came to know him as a person with a tireless capacity for work and absolute exactness, coupled with a sense for fairness, accessibility and courtesy. After his retirement, he showed continued interest in and affinity with the activities of the Bureau and the Language Committee. Colleagues can bear testimony that Niël was, as is said in Yiddish, a 'mensch': a selfless person characterized by integrity, genuineness and compassion. He spoke and wrote Afrikaans in such a way that it demonstrated its validity as a medium in all language situations. For this devotion and for the person that he was, his memory is honoured.

An Analysis of *The Oxford Guide to Practical Lexicography* (Atkins and Rundell 2008)

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

Abstract: Since at least a decade ago, the lexicographic community at large has been demanding that a modern textbook be designed — one that would place corpora at the centre of the lexicographic enterprise. Written by two of the most respected practising lexicographers, this book has finally arrived, and delivers on very many levels. This review article presents a critical analysis of its features.

Keywords: LEXICOGRAPHY, LEARNERS' DICTIONARY, MONOLINGUAL, BILINGUAL, CORPUS, FRAME SEMANTICS, ENGLISH, FRENCH, TEXTBOOK

Samenvatting: Een analyse van *The Oxford Guide to Practical Lexicography* (Atkins en Rundell 2008). Al minstens tien jaar lang eist de volledige lexicografische gemeenschap dat een modern tekstboek zou worden ontworpen — één dat corpora in het centrum van de lexicografische belangstelling zou plaatsen. Geschreven door twee van de meest gerespecteerde praktiserende lexicografen, is dit boek er nu eindelijk, en het ontgoochelt niet. Dit recensieartikel analyseert de kenmerken ervan kritisch.

Slutelwoorde: LEXICOGRAFIE, LEERWOORDENBOEK, VERKLAREND (MONOLINGUAAL), VERTALEND (BILINGUAAL), CORPUS, FRAME SEMANTICS, ENGELS, FRANS, TEKSTBOEK

What matters — and this is critical — is not the writer's intention but the reader's interpretation. (Atkins and Rundell 2008: 439)

1. Striking gold

Lexicographers worldwide rejoice — the definite textbook on how to make a dictionary from a corpus has arrived! It is called *The Oxford Guide to Practical Lexicography*, for short OGPL, and was written by B.T. Sue Atkins and Michael Rundell. During the book launch on 17 July 2008, at the Thirteenth EURALEX International Congress in Barcelona, John Davey, Linguistics Editor at OUP UK, simply stated that he had 'struck gold'. He could not have summarized it any better.

In the first edition of the *Macmillan English Dictionary for Advanced Learners*, MED-1 (2002), 'strike gold' is sense 13a under **strike**¹ *verb*. In the second edition, MED-2 (2007), 'strike gold' is listed in a separate phrase block under the same headword. In the latest *Collins-Robert French Dictionary*, CRFD-8 (2006), one reads under sense 2 *transitive verb*, subsense c: '**to strike gold** (figurative use): trouver le filon*' (where the asterisk marks that the French equivalent, but not the English, is informal). One and the same expression, three (very) different treatments. Is one to conclude — right away — that lexicography is not an exact science, that there are no theories that govern it? Not at all: About every page of OGPL reconfirms both the highly scientific nature of the lexicographic enterprise and the need for strong theoretical underpinnings. Theories evolve, users become ever more demanding, the types of dictionaries spiral outward. Variation and continual improvement, therefore, are inherent. Although reference works have been produced for thousands of years, OGPL proves that the field of lexicography is more alive than ever.

2. Importing, exporting and reimporting gold

With reference to the line of dictionaries cited in Section 1, Sue Atkins was the General Editor of the first *Collins-Robert English-French Dictionary* (Atkins and Robert 1978). Michael Rundell has been the Editor-in-Chief of the Macmillan family of learners' dictionaries since their inception. Both were involved in numerous other dictionary projects, beginning in 1966 for Sue, in 1980 for Michael, including the revolutionary COBUILD project of the 1980s.

To many readers of *Lexikos*, Sue and Michael are of course well known. Back in September 1997, thus just two years after the establishment of PanSALB, and in order to prepare the future lexicographers of the 'soon-to-be established' eleven National Lexicography Units (a process finalized in March 2001), Sue Atkins and Michael Rundell, joined by Edmund Weiner, presented the two-week-long (monolingual) SALEX Lexicographical Training Course. This course was organized by Penny Silva, then Executive Director of the Dictionary Unit for South African English in Grahamstown, South Africa, currently Director of Editorial Projects for the *Oxford English Dictionary* in Oxford, United Kingdom. One year later, in September 1998, Sue and Michael returned to South Africa, for the first week of the follow-up (bilingual) AFRILEX-SALEX Lexicographical Training Course, organized by AFRILEX at the University of Pretoria.

While both Sue and Michael had already been involved in teaching courses for publishing houses for about a decade, the magnitude of the South African experience (with 35 participants at SALEX and 50 at AFRILEX-SALEX) surely contributed to their decision to turn their efforts into an annual event. Indeed, three years later, and joined by the computational linguist Adam Kilgarriff, they began their LEXICOM series, an intensive five-day workshop in lexicography and lexical computing. The detailed notes drawn up for the South

African courses, and perfected and complemented over the years for the LEXI-COM courses, eventually 'morphed into' (p. ix) OGPL.

The present commentator has been in the fortunate position to be a fellow traveller for a decade now, with attendances of the 1998 (Pretoria) and 2001 (Brighton) courses as a participant, and the 2005 (Brno), 2006 (Opatija), 2007 (Brno) and 2008 (Barcelona) courses as a guest presenter of the dictionary writing system *TshwaneLex*. One thing is certain: Sue and Michael kept the absolute best from their course notes, and added even better material on top.

OGPL, then, clearly has some of its roots in South Africa. As if to celebrate this fact in its own way, it is undoubtedly befitting that the University of Pretoria conferred an honorary doctorate on Sue Atkins on 5 September 2008 (see pp. 412-416 in this issue of *Lexikos*).

3. A gold and diamond-studded trophy

The main **strength** of OGPL is that it has a crystal-clear goal, and that the authors have managed to stay focused on that for over 550 pages. That goal is a detailed description of how today's monolingual and bilingual *corpus-driven commercial learners' dictionaries* are compiled. Although most reference works are actually adaptations of existing ones, Atkins and Rundell have decided to go all the way, and present every single step 'from the first twinkle in the eye of the publisher' (p. 18) through to the completion of the central dictionary text. Given the background of the authors, it will not surprise that the description of the compilation of a monolingual learners' dictionary is for an *English* one, and that that of a bilingual learners' dictionary is for an *English to French* one (and only that direction). Atkins and Rundell also decided to focus on the production of *paper* dictionaries for *human* users. Electronic dictionaries and NLP applications are mentioned where appropriate only.

The main **novelty** of OGPL is the minute description (which runs through and frames the entire book) of the so-called 'two- and threefold lexicographic process' — for monolingual and bilingual dictionaries respectively — used to compile corpus-driven dictionaries from scratch. Ironically, about this Atkins and Rundell state: 'we make no special claim for the methodology we outline in this book [...] What we describe here is what has worked well for us over a number of years' (p. 2). On the contrary, the whole point of the book is to make a strong case for exactly that approach.

Following an introductory chapter, the book is divided into **three parts**: the first deals with 'Pre-lexicography' (six chapters), the second with 'Analysing the data' (two chapters), and the third with 'Compiling the entry' (three chapters). This in itself is an extremely bold move. Recall that the aim was to describe the making of a dictionary from a corpus, yet the first part (i.e. before any analysis is being done) runs over an astonishing 250 pages. Rather than start with fireworks, and instead of discussing one topic at a time, the authors opted for a multilayered approach, whereby the same topic is revisited time

and again as the argument progresses. This results in a very large number of forward and backward cross-references. In §7.2.7.2, for example, one finds the following **forward cross-references** to sections in Chapters 9, 10 and 12: 'Choosing examples is a very important part of entry-writing, and is further discussed in §9.2.4, §10.8 and §12.3.3' (p. 225). Or, for the reverse, in §12.2.5 the following **backward cross-references** are found to sections in Chapters 6, 7 and 9: 'The vocabulary types represented by the linguistic labels are introduced in §6.4.1.4; the way the labels function is explained in §7.2.8; and their use in the database is set out in §9.2.9' (p. 496). There is thus a real need to read everything and it is hard to lift out sections, as one may miss a point (or two, three).

In **Chapter 1**, the 'Introduction', the book and its topic are briefly introduced. It can also be read as a 'thank you/acknowledgement' section. From the first sentence onwards, it is stressed that OGPL is about 'creating a dictionary' (p. 1), and by implication thus not about publishing, testing, analyzing, theorizing, etc. An important point is the following: 'You learn about lexicography by doing it, by training other people to do it [...], and by talking about it with colleagues' (p. 9). To any aspiring lexicographers: OGPL can thus best be used when put into practice.

Part I, 'Pre-lexicography', is the longest part with six chapters. The first of those is **Chapter 2** on 'Dictionary types and dictionary users'. Following an introductory section in which the entire process line of dictionary production is sketched, this chapter deals with the types of dictionary, the types of dictionary user, and then, combining the two, a section on tailoring the entry to the user who needs it. §2.4.1, on how to do the latter for monolingual dictionaries, is a nice, gentle analysis within the framework of this chapter's title. From the equivalent for bilingual dictionaries, one will do well to remember the following right from the start: 'the bilingual dictionary is more complex, and less amenable to clear explanations, than all but the most scholarly and sophisticated of the monolinguals' (p. 43).

The topic of **Chapter 3** is 'Lexicographic evidence', or reformulated: How to make sure that your dictionary is reliable? Here, *introspection* (with the required snipe at Noam Chomsky on p. 49) and *informant-testing* are contrasted with the use of *citations* and *reading programmes*. Although each of those is valuable, the type of evidence this book puts first and foremost is of course that found in large collections of electronic texts, commonly known as *corpora*. OGPL covers their design and the data collection in detail, and also sketches processing and annotation issues. The coverage is informative at all times, full of excellent advice, as in: 'never offer to pay for permission to include a text, and never agree to such a request from a copyright holder. Once money starts changing hands (even if for a single text in a single corpus), a precedent would be established that could have fatal consequences to corpus-creation efforts worldwide' (p. 83).

Chapter 4, titled 'Methods and resources', is a mixed bag. In the middle of this chapter the two chief 'resources' are discussed: a *corpus query system* (CQS) on the one hand, and *dictionary writing software* (DWS) on the other. The CQS

that is singled out is the Sketch Engine, with its linked lexical profiling tool, the Word Sketch. Regarding DWSs, Atkins and Rundell write: 'Two of the best and most widely used packages are IDM's Dictionary Production System [...] and the TshwaneLex dictionary compilation software' (p. 114). The last two sections of this chapter look into the *Style Guide* and the use of *template entries* (i.e. lexical-set compiling): 'As a rule, the members of a lexical set pose the same kinds of lexicographic problem and should be handled in the same way in a dictionary' (p. 124). The section on template entries is exceptionally well executed, and is one of the highlights of OGPL.

The opening section of this chapter focuses on the 'method' advocated by the authors. This method is revolutionary, as the compilation route is not straight from corpus to dictionary, but rather, a *twofold lexicographic process* whereby a database editor will first analyse the corpus and populate a pre-dictionary database. In a second step, a dictionary editor will then synthesize those findings in order to produce a dictionary. This is the process for monolingual dictionaries. For bilingual dictionaries, an extra step is needed in-between, hence the characterization as a *threefold lexicographic process*, whereby a translator populates the pre-dictionary database with (selected) translations. Without a good understanding of this process, OGPL cannot be used successfully: Part II (Chapters 8 and 9) is entirely devoted to the first step ('analysis'), while in Part III, Chapter 10 is devoted to the second step ('synthesis') for monolingual dictionaries, and Chapters 11 and 12 to the second and third steps ('transfer' and 'synthesis') for bilingual dictionaries.

In **Chapter 5**, 'Linguistic theory meets lexicography', the sense relationships *hyponymy*, *synonymy*, *meronymy* and *regular polysemy* are contrasted with *complementary antonymy*, *polar antonymy*, *directional antonymy* and *converseness*. The central section in this chapter, however, is the presentation of *Frame Semantics*, about which the authors say: 'the application of this theory to practical lexicography results in [an] approach to lexicographic relevance [...] which helps lexicographers to identify useful facts in corpus texts' (p. 144). In other words, Frame Semantics is the core linguistic theory which is advanced in OGPL to analyze corpus data. The last section of the chapter is on *lexicographic relevance*, which is actually an introduction to Chapter 9.

In a more traditional framework, **Chapter 6**, on 'Planning the dictionary', would be referred to with *macrostructure* decisions. The aims are: 'in §6.1 to clarify some of the basic terms and concepts; in §6.2 to name and describe the different kinds of words and phrases which you have to be able to recognize in corpus data and which can be dealt with as headwords in a dictionary; in §6.3 to set out the various large components of a print dictionary (the actual dictionary text and other material); in §6.4 to explain features you have to consider when deciding the words to include in the dictionary; in §6.5 to look at the other main decisions that have to be made about the headword list; and in §6.6 to describe the principal types of entry that are to be found in most current dictionaries' (p. 161). In the process, a large amount of (new) vocabulary is introduced and defined. The basic terms and concepts are *tokens*, *types*, *lemmas*, *lexi-*

cal units (LUs), and *headwords* (§6.1). The types of lexical items discussed include: *lexical words* (nouns, adjectives, verbs, adverbs, interjections); *grammatical words* (prepositions, conjunctions, pronouns, auxiliary verbs, determiners); *alphabetisms, acronyms, contractions; bound affixes, productive affixes, combining forms; fixed and semi-fixed phrases* (transparent collocations, fixed phrases, similes, catch phrases, proverbs, quotations, greetings, phatic phrases); *phrasal idioms; figurative, semi-figurative and functional compounds; phrasal verbs; and support verb constructions* (§6.2). The constituent parts of a dictionary are of course: *front, mid- and back matter*, plus the *A–Z entries* (§6.3). The headword list proper consists of: *common words* (for which one needs to consider wordclass, lexical form, lexical structure, and vocabulary types (which leads to linguistic labelling)) and *proper names* (place names, personal names, and other names) (§6.4). In order to organize the headword list one needs to know more about: *alphabetization, syllabification, and homographs* (§6.5). Lastly, the types of entry are: *standard lexical, abbreviation, grammatical, and encyclopedic* (§6.6). The above enumeration serves as an example to show the dedication of the authors: every possible term and concept that is needed is also covered.

In a more traditional framework, **Chapter 7**, on 'Planning the entry', would be referred to with *microstructure* decisions. In analogy with Chapter 6, a very large number of terms and concepts are presented, all of them adequately illustrated by means of marked-up components of dictionary excerpts. In the section on labelling, Ogden and Richards's 'meaning triangle' is a delight. Tucked away in this chapter's back pocket, namely in subsection §7.2.11, is a basic introduction to electronic dictionaries.

In **Part II**, 'Analysing the data', Atkins and Rundell 'get down to business' (p. 261). They do this in two phases, first focusing on the identification of word senses of polysemous lemmas in Chapter 8, and then fleshing out those senses (or more correctly *lexical units*) in Chapter 9. As pointed out above, at this point no dictionary is being produced: a language (*in casu* English) is analyzed in detail, and the results of that (corpus) analysis are stored in a pre-dictionary database.

Chapter 8, titled 'Building the database (1): word senses' is a core chapter, which is no doubt why it was singled out to be offered as a free sample of OGPL on the publisher's website. The task can be described as follows: 'Meanings exist in infinite numbers of discrete communicative events, while the senses in a dictionary represent lexicographers' attempts to impose some order on this babel' (p. 311). Of the various 'theories' offered, one is especially relevant: *prototype theory* — which is complemented by discussions of *polysemy* and *homonymy, lexical semantics* and *motivated polysemy* (i.e. specialization, regular polysemy, figurative uses, metaphorical sets, conceptual metaphor, and metonymy), as well as a string of other useful ideas. Despite of all this, the authors point out: 'Learning about these ideas won't necessarily make the process of identifying word senses any easier, but you will tackle the job with greater confidence if you understand the underlying systems, and you will be better equipped to make good judgments in the more marginal, less clear-cut cases'

(p. 294). OGPL certainly offers a wide array of practical strategies for successful *word sense disambiguation* (WSD), strategies which will hopefully bring lumpers and splitters closer to one another — at least at the analysis (pre-dictionary) stage.

The bulk of **Chapter 9**, 'Building the database (2): the lexical unit', deals with *grammar in the pre-dictionary database* (§9.2.5), with a focus on (English) verbs, nouns, adjectives, and adverbs. For each of these wordclasses, the lexicographically relevant co-constituents are first listed in detail, then it is shown how those constructions can be retrieved from the corpus, and finally it is illustrated how examples may be selected to support each instance. Those four subsections are followed by what must be the absolute highlight of OGPL, namely a detailed ten-page-long 'case study on using complements in sense analysis, and in recording them' (p. 349). The headword being studied is the verb 'cook', and unlike the four 'theoretical' descriptions which preceded this subsection, the correct order is used here, viz. (1) analysing the corpus lines, (2) identifying the constructions to be recorded, and (3) recording complementation in the database. Note thus the order of (1) and (2). The beauty of this case study is that it is clearly shown how analyzing concordances in Frame Semantics 'Light' terms gives the valency patterns of each lexical unit of the headword. A pity indeed, then, that this gem is tucked away deep into this chapter (under §9.2.5.5) and that no cross-references whatsoever to it are to be found anywhere. It is not even listed under the entry 'frame semantics' in the book's Index. The rest of this chapter basically tells the reader how other database fields (such as headword, wordclass, meaning, examples, multiword expressions, collocates, patterns, labels, cross-references, and comments) are dealt with.

Part III, 'Compiling the entry', is 'the moment of truth' (p. 383). This section is perhaps best introduced with two quotations from Chapter 9: 'A database entry should be at least two or three times bigger than the final dictionary entry' (p. 322). 'Take the case of the word *coffee*. This word, in the sense of the substance, either solid or liquid, is uncountable; however when it is used to denote a single unit of coffee (in a cup, for example) it becomes countable; a similar phenomenon occurs when the word is used to mean 'type of coffee'. These three senses would normally be considered as different LUs in the database: what you do with them in a dictionary depends (as always) on the type of dictionary' (pp. 337-338). Dictionary editors will thus study all the facts that have been recorded by the database editors, and 'distil' from that exactly that type of dictionary which is deemed appropriate for the envisaged target users. A new team of dictionary editors may resynthesize the facts for a new edition in a slightly different way (which may also be accompanied by changes to the Style Guide), yet another team may extract a very different type of reference work altogether from the same database, whilst the largest ramifications will of course be seen when a second language is added to the database in order to produce a bilingual dictionary. Recall in this regard, for example, the various treatments of 'strike gold' in Section 1.

Chapter 10, 'Building the monolingual entry', concludes the discussion of the creation of a monolingual learners' dictionary. While most information during this stage may 'simply' be selected from the pre-dictionary database ('synthesis'), two new elements surface: presentation and definition writing. The first involves decisions on how to distribute the information, or thus what to do with, and where to place multiword expressions, run-ons, dictionary senses, grammar and labels. The second is a truly new task, which is why the bulk of this chapter (a massive fifty pages, or about 10% of OGPL) is concerned with how to write definitions that match the corpus-derived facts. Examples that had already been collected may at that point be selected from the database to support and complement the newly written definitions. All along, a close eye is also kept on the user profile and the Style Guide, while template entries are used whenever appropriate to ensure consistency.

Chapter 11, 'The translation stage', deals with stage two ('transfer'), of the threefold lexicographic process which characterizes the compilation of a bilingual dictionary. Here, translators 'work through each LU (both word senses and multiword expressions), adding target-language (TL) translations, going forward and back over the entry, and seeing which TL word seems to fit best as the first, or 'direct', translation — in essence, the word that suits most of the contexts before them. Then they decide which of the remaining contexts (those which the direct translation doesn't fit) are important enough to be kept in the entry, and translate the headword in these contexts' (p. 465). Important concepts that come into play here are *context-sensitive* vs. *context-free* translations (the latter is the aim of the dictionary), and *denotation* (literal meaning) vs. *connotation* (figurative meaning).

Chapter 12, 'Building the bilingual entry', mirrors Chapter 10 in more than one respect. Here, the bilingual dictionary editor will edit the entries from the translated pre-dictionary database, completing stage three ('synthesis') of the threefold bilingual process. The crucial difference with the monolingual effort — at least within the framework advocated by OGPL — is that 'dictionary senses' in a bilingual are not really senses of the headword at all, but simply the most user-friendly way to structure the material' (p. 500). As if to shatter one's beliefs even more, OGPL further claims that: 'It is naïve to think that you can lift stretches of corpus text, translate them, and produce really useful examples' (p. 507). After over five hundred pages of excited build-up, one ends up throwing out both dictionary senses and real corpus examples. All good things must come to an end.

4. Technical aspects

If anything, Section 3 demonstrates the enormous ground covered by OGPL. Inevitably, such a mammoth undertaking also resulted in some infelicities. Those are looked into below.

4.1 On the register and background

To those of us used to a strict academic **register**, the informal style of OGPL is immediately noticeable. Not only do very many sentences start with 'And', 'But', or 'So', large sections truly read like a novel, while the reader is often (but inconsistently) addressed. Atkins and Rundell use 'easy' language throughout, shunning any unnecessary metalanguage that is all too common among those who think that there is such a thing as 'theoretical lexicography'. Every now and then a 'hard word' is dropped, however, as the selection in Addendum 1 shows.

Written in English, the primary language of study for the monolingual effort is also English (although a single Spanish definition is flashed, p. 440). French is added to the equation for the bilingual effort (with one case of Finnish however, p. 41). OGPL, then, is fully embedded into the English and European cultural world, so mentions of say Westminster (p. 273), or the contributions of the Greek Aristotle (p. 437) and the Swede Linnaeus (p. 437), are all taken for granted. Specifically for lexicography, it is assumed that every reader already knows who Samuel Johnson (pp. 5, 48, etc.), James Murray (pp. 50, 52, 275), Noah Webster (p. 52), or A.S. Hornby (p. 455 (and p. 449)) are and what they did. Similarly for some lexicographic resources: the BNC is first mentioned on p. 41 (without any forward cross-reference to what it is on pp. 58-61); WordNet is only 'dropped' (p. 82). Readers also need some level of linguistic background to understand conventions such as '+HUMAN' (p. 155), the meaning of terms such as 'attributive', 'predicative', or (especially relevant in French) 'subjunctive' (all p. 221), or say 'Grice's maxim of quantity' (p. 456). All of these are not necessarily problematic; it depends on the audience (and the **background** of that audience) one wishes to reach.

4.2 On the general structure

As is clear from Section 3 above, the overall structure, with the main chapters distributed across three parts, is very effective. Early on in each of the main chapters one also finds a **schematic overview** summarizing that chapter's contents and structure. This is an excellent idea (although the 'Preliminaries' are missing in Figg. 7.1 and 8.1 (>< Fig. 6.1) while Fig. 10.1 is not referred to).

Unfortunately, in several places the **hierarchy** (and the related numbering system) went wrong. For instance, §2.1.2 should have been §2.1.1.2; §7.2.8.8 should not have been on the same level as the seven previous subsections; §8.3.4-6 should have been §8.3.3.1-3; §8.3.7 should have been §8.4; the current §8.4 then changes to §8.5 (with the current heading at §8.5 best dropped). Related, sections called 'Conclusions' are sometimes not conclusions at all, as is for example the case with §8.6 and all its subsections: this reads like an add-on, with far too much new information for it to be a conclusion. The same is true for the material on p. 431, which does not belong to the conclusion of §10.5.8.

Perhaps mimicking the latest range of learners' dictionaries, OGPL is printed in two **colours**: black (as well as some grey tints) and blue. This is especially effective when **dictionary sections** are shaded for ease of reference. The entire background colour of the so-called 'boxes' is also blue. **Boxes** are used to provide background information, and (in Chapter 9) for core tables and discussion points. Unfortunately, some of the boxes are not referred to (namely Boxes 3.1, 4.1, 7.1, 7.2, 7.3 and 9.7). While some of the extra information in such boxes is indeed truly excellent (e.g. Boxes 4.2 and 10.1), at times the material seems to be close to meaningless (e.g. Box 4.3). Also set in blue, is the right-pointing arrow (➔) which precedes all so-called 'rule of thumb practical tips' (p. 229).

Another excellent feature are the various lists of **pros and cons** to help readers make decisions for themselves, such as the enumeration of the advantages and disadvantages of citations (pp. 51-52), the pros and cons that deal with whether the 'first cut' in microstructure design should be based on grammar or meaning (pp. 248-249), or the options relating to multiword expressions as headwords (pp. 254-255). In most cases, however, the options are not made explicit and need to be 'derived' from the arguments presented, as in the insightful discussion of the use of full sentence definitions (pp. 441-443), or the presentation of the source and characteristics of good examples (pp. 455-461).

A true textbook feature comes in the form of **exercises**. While every chapter presents those at the end, they are missing from Chapters 1 (of course), 3, 4 (but the last section may be seen as one), and 5. The exercises are mostly simply straightforward applications of the points that were covered. Each chapter is formally concluded with a **Reading** section (for which, see Section 5.2 below).

4.3 On the multilayered build-up

As pointed out in Section 3 above, the presentation of the two- and threefold lexicographic process automatically led to a multilayered build-up of the entire line of argument. While this often works out well, there are some annoying consequences. Chief among those are cases of unnecessary **repetition**. The concept of a 'lexical set', for example, is first defined on p. 123, then again on p. 139, and yet again on p. 490. Large swathes of Chapter 12 are an especially bad example: pp. 484-488 are a near-calque of pp. 385-389 a hundred pages earlier. Likewise, §12.1.2 mimics §10.1.2, and §12.1.3 mimics §10.1.3. Surely, a simple cross-reference together with a summary of the differences would have sufficed (as was resorted to in the last paragraph of §12.2.1). Repetition sometimes also occurs within a single chapter, as for the two 'bullets' as well as the 'valency patterns' on p. 148 and again on p. 155. Some types of repetition result in a time warp, as in 'a day or two' (p. 508) becoming 'a week or two' (p. 513).

The **order** of the introduction of certain terms and lexicographic devices is not always correct either. For example, although the (undefined) concept 'lexical unit (LU)' had already been used a number of times, it suddenly takes central stage in the discussion of template entries (pp. 124 *fn*), but one only finds a

(first) short explanation in footnote 2 on p. 131, and a (second) full explanation on pp. 162-163. Of course, it is not always easy or even possible to introduce one aspect at the time, but the way the term 'canonical form' was handled on p. 168 — in a brief footnote that includes a cross-reference to the fuller explanation — is certainly more effective for a textbook. A similar reasoning can be made for the term 'lemma', which is used on p. 150, but only defined on p. 162, or for the term 'itemizer' on p. 323, defined on p. 371.

The same holds for some **symbols** and **metalinguage** in quoted dictionary articles. Two examples: The function of the (gender) letters '*m*', '*f*', '*fpl*', etc. should have been explained from their first appearance (in Fig. 6.14, p. 194) onwards. The (informality) markers '*' and '**' should also have been explained from their first appearance (in Fig. 7.6, p. 212) onwards, rather than (only partly) on p. 232, and (fully) on p. 498.

While some **building blocks** of the argument are indeed nicely stacked on top of one another (compare for instance Fig. 5.25 with Fig. 8.8), some blocks do not seem to fit. §9.3, following a sixty-page-long §9.2, is just 0.3 pages long, and amounts to one backward and two forward cross-references. Likewise, §12.2.3 seems to be just a place holder for §12.3.1. At other times, blocks, such as the 'dictionary pilot study' (p. 196), appear unexpectedly, once, and are never returned to.

5. References and cross-references

References and cross-references come in many different forms in OGPL. In the front matter, one finds a list of abbreviations which refer to dictionaries, and another one for an assortment of concepts; each chapter is concluded with a Reading section; the main text includes classical references to the Bibliography while the different building blocks cross-refer to one another; and in the back matter one finds a Bibliography plus an Index. References and cross-references hold everything together; unfortunately, they are sometimes a bit inconsistent. This is discussed below.

5.1 Abbreviations

On p. x, a list of **dictionary abbreviations** is presented. All of them are indeed used in OGPL, but, given there seems to be only one instance of CALD (on p. 389) one could as well have done without that dictionary abbreviation. The abbreviations are, however, not always used, and are sometimes written out in full: OALD on p. 22, MWC on p. 310, MW-3 on pp. 312, 417, 426, ODE on p. 400. The system used to refer to these dictionaries is not consistent either (e.g. '*CRFD-1998*' in Fig. 7.6 vs. '*CRFD-5 (1998)*' in Fig. 7.7). Quite surprisingly, not a single of these dictionaries has been provided with full bibliographic details. The same is true for the many other dictionaries, written out in full, that

are referred to in the text. Not one made it into the Bibliography.

On pp. xi and xii a list of '**other abbreviations**' is presented. One of them does not seem to be used anywhere in the text (VP), another one was already included with the dictionaries (OED), while quite a number are missing (L1, L2, N, XML). Several are in any case written out in full in the text (e.g. IPA on pp. 29 and 37, MLD on p. 410 n7, DWS on pp. 358, 489, 495 and 497).

5.2 Reading section

At the end of each chapter, a list of references is included to relevant background material. That Reading section is always divided into '**Recommended reading**' (R) and '**Further reading on related topics**' (F), sometimes complemented by a section '**Websites**' (W).¹ Quite surprisingly, for some Reading sections not a single source has been referred to from the preceding text (Chapters 7, 11, 12); for others just a handful (Chapters 9, 10-F). Two sources appear in both the R- and F-list, which is a contradiction (Rundell 2002 on p. 316, and Hanks 1987 on p. 464). One is missing from the Bibliography (Melčuk [sic] 2000 on p. 316).

5.3 Main text

If I have seen further it is by standing on ye shoulders of Giants. (Isaac Newton, Letter to Robert Hooke, 15 February 1676)

In order to be able to write such a revolutionary textbook as OGPL, Atkins and Rundell had to build on earlier work. At times, however, the text cries out for references. Proper and appropriate references to such works would not have diminished their own standing. Some examples: to the use of computers in the dictionary-making process, on pp. 3 and 112: **Urdang 1966, 1984**; to the published proceedings of the First Fulbright Colloquium, on p. 32: **Ibson 1986**; to *CLAWS*, on p. 90; to an analysis of the 'synonyms' shut and close, on pp. 134-135: **Lewandowska-Tomaszczyk and Hanks 1996**; to semantic prosody, on p. 304 (although that term is not mentioned there, only on pp. 321, 375-376): **Sinclair 1987, 1998**; to the Theory of Norms and Exploitations, on pp. 309, 312, 397: **Hanks 2004**; or to 'the probable and the possible', on pp. 314, 398: **Hanks 2001**. The authors could have referred more often to their own work as well. To give one example: The gem described in Section 3 (Chapter 9) above, is based on **Atkins 2002**.

While it is understandable that **references** to novels (pp. 47, 73 n22, 270, 297, 408) have not been included in the Bibliography, and while it is true that one can simply google for *Google* (p. 48) or *Wiktionary* and *Wikipedia* (both p. 422), it would have been helpful if the source of all the concordances (or the name of the corpora used) had been provided throughout (where do, for instance, the examples on pp. 270, 279, 295, 297, etc. come from?). Or, again,

should one always simply resort to *Google* to find the source of a spoken transcript such as the one on p. 78, or to see who is being quoted in the 'Consignia anecdote' on p. 375? (The sources are CNN and *Industry News* respectively.) Saying that a quotation is from Johnson is also not good enough: those on pp. 5, 9 and 178 are from Johnson (1747), while the one on p. 448 is from Johnson (1755). Similarly for the quotation of Fillmore on p. 410; it is from Fillmore (2003).²

The different building blocks of the main text are held together by literally hundreds and hundreds of forward and backward **cross-references**. Most of these cross-references were placed in the running text itself, others have been relegated to footnotes for no obvious reason. Even so, some sections cry for more cross-references (so, at the bottom of p. 208 for instance, one would have liked to see a mention of §10.6.3.1).³

5.4 Bibliography

The Bibliography does *not* bring together all the bibliographic details of the material referred to in OGPL. One reason was given in the paragraph on 'references' in the previous Section 5.3. Another reason is that references that are only mentioned in footnotes and not in the running text have been provided with full bibliographic details in those very footnotes (pp. 48 n3, 52 n5, 55 n10, 56 n12, 74 n24, 226 n13, 293 n20 (2 x), 369 n16, 401 n5, 419 n11, 422 n14, 436 n22).⁴ Full bibliographic details may also, and only, be found in boxes (Box 3.1) and even in the running text (p. 400). It is not immediately clear why this user-unfriendly option was chosen.

5.5 Index

A ten-page-long Index concludes OGPL. Regrettably, it is a subject index only; there is no author index. All the subjects and page numbers covered indeed appear in OGPL, but the coverage seems far from exhaustive. Firstly, numerous concepts have not been included at all. A random sample follows, including selected page numbers: active 506; faux amis 475; *pertainim* 305, 339, 343, 510; semantic prosody (304,) 321, 375-376; stratified sampling 64, 66, 68; passive 506.⁵

Secondly, often one or more page numbers are missing for those terms that are covered. A small sample: back matter +506; Brown Corpus +80; connotation +468-469; denotation +469; electronic dictionary +477; frame semantics +349 *fn* (esp. 354-355); IPA +29; priming +270, +297. The latter, 'priming', is an interesting case. On p. 270 and again on p. 297, the term is used without being defined, nor with any forward cross-references to where it is eventually defined (on p. 307). While it is acceptable not to give the meaning away too early, in order to 'prime' the readers, all instances *should* have been covered in the Index nonetheless.

The third type of problem with the Index is that some terms are treated in an idiosyncratic way. One example will suffice. While 'definiendum' (used on pp. 421, 440 and 435) is not included in the Index, 'definienda (form or function of) 439-440' is. It is not clear why the plural form was used as an entry for the Index, especially that it does not appear as a plural on pp. 439-440.

Lastly, the Index could have been put to good use in order to somewhat reduce the very large number of references and cross-references throughout the text, and especially in footnotes.

6. Errors and suggestions for improvement

Errors are inevitable in any project, so too in OGPL. Addendum 2 brings together around seventy (types of) errors. If the comprehensiveness of OGPL is taken into consideration, that number is not so dramatic, yet given all of them could and should have been picked up by the publisher, they remain regrettable. In Addendum 2, the errors were classified according to whether they are general or specific. In the second group, errors in quoted dictionary text, concordances, graphs and boxes are looked into. The first group presents the remaining typos, as well as problems with inflection and syntax, capitalization, and punctuation, and also lists inconsistencies, font and layout issues.

Addendum 3 contains some suggestions for improvements, again subdivided into subsections: more conventional formulations, likely better options, sections to improve upon, incorrect statements, and additional (recent) references.

7. Discussion points

There can be no question that for this type of textbook, the two **best** candidates joined forces. B.T. Sue Atkins, with her background of working as a lexicographer, manager and consultant for the great publishing houses of Collins/Robert and Oxford/Hachette, attended to the bilingual component, while Michael Rundell, with his experience of working as a manager for Longman and Macmillan, was responsible for the monolingual component. During their careers they certainly also came into contact with the best computational lexicographers. OGPL, then, was bound to be a hit. Their strength, however, also encapsulates some weaknesses.

Obviously, not 'everything' is (nor can it be) covered in OGPL. Very little if anything is said about **management** and time lines, for example, and tools to help with that.⁶ Open any of the modern learners' dictionaries, on any page, and one wonders how long each of those articles took to be compiled. Or more down to earth: Fig. 9.10 summarizes just one LU of the verb *watch*, so one shudders at the thought of doing this type of analysis for every single sense of every single word of a language. On how to handle this, the authors could have

given us more information, as they have had to deal with it time and again.

Specifically for bilingual lexicography, the reader may also wonder how the **reverse side** should be compiled. Is one to start afresh, with a brand-new analysis of the second language, or could one hope to 'reverse' and thus reuse large sections of the side already compiled? Plus, in either case, how is one to cope with the 'reversibility principle'?⁷ Again, some coverage of this would have been welcome.

Dictionaries of the type being described easily require over a hundred **contributors**, with many more collaborators in addition to 'database editor', 'translator' and 'dictionary editor', as any dictionary title page will confirm. How is one to organize this? What is best outsourced (function words, subject-specific terms, illustrations, grammatical sketches, all extra matter, etc.)? As the textbook stands, one may end up with a perfect central section, without any paper to wrap it in.

One drawback to the near-exhaustive presentation of the creation of a dictionary's central section is that the process intimidates. Surely, one knows all the secrets to become a lexicographer now, but also knows one cannot possibly assemble the materials. For the sake of the lesser fortunate, the options should (and could) have been downplayed every now and then. With regard to **corpora** and a CQS, why was the obvious not even mentioned, namely to simply dump plain text files into say *WordSmith*? Conversely, true corpus building for many (non-English) languages involves more than is covered by OGPL. For those without much or even any digital data, for example, scanning of printed data requires smart *optical character recognition* (OCR) 'tricks'. (While making a corpus from an oral language requires even more; but we would not explore this.) OGPL's near-exclusive focus on developments in and around the British Isles also failed to take account of some important work done elsewhere. Still on the issue of corpora, p. 79 boasts, with emphasis added: 'The New Corpus for Ireland — collected in 2003–4 — is an *early example* of a corpus which includes web data.' Hardly impressive, one would say, as in mid-October 1999, the present commentator had already built a corpus of Swahili sourced entirely from the Web (cf. e.g. De Schryver 2002: 268).

With regard to the selection of **examples** to populate the pre-dictionary database, OGPL states as ideal, with numbering added: '(1) It's short. (2) It provides an 'informative context' for the headword [...] (3) It has no words in it that are more difficult to understand than the headword. (4) It doesn't include words at variance with the register, style, region, etc. of the headword. (5) It doesn't contain the name of a real person, living or dead. [...] (6) It has no foreign words in it' (p. 330). If one compares this with the various excerpts of the pre-dictionary database in OGPL, one notices trespasses against more or less every point. Examples 1a and 2a in Fig. 9.13 for example are *very* long, while example 6b in the same Fig. 9.13 mentions Yeltsin. Compare this with (1) and (5). It seems as if OGPL here mixes the pre-dictionary database with the final dictionary.

Cases such as these may actually be a sign that the dictionary-writing process known as the **'two- and threefold lexicographic process'** — as *described*, and *with* the suggested software — has actually never been implemented. This is a strong statement, but one based on an impression that runs right through the reading of OGPL. If one goes back to the software that is being presented to support the dictionary-writing process (p. 103), one sees that two types are offered: a CQS and a DWS. Instantly, one not only wonders how data from the CQS ends up in the DWS (the 'dictionary database'), but more importantly, one needs to know where the 'pre-dictionary database' (which stores the analysis) fits in. The ensuing discussions of CQS (§4.3.1) and DWS (§4.3.2) do not give any hint. Each subsequent mention of 'database' in OGPL was carefully scrutinized (amongst others on pp. 123, 126, 157, 174, 264, 268, 317, 318, 319, 336), but to no avail. On p. 319 for example, one reads: 'Database and dictionary differ in structure as well as size', so one wonders which tools 'transform' that structure. On p. 336 a database extract is even shown (as text in a word-processor table), but one wonders why that extract is not rather a screenshot, as is customary when demonstrating software features. The one and only 'clue' is found on p. 358, where the 'pre-dictionary database' being populated seems to be equated with the 'dictionary writing system'. Each next mention of 'database' (such as on pp. 470 or 480 *fn*, both ideal candidates for screenshots) remains vague on this issue. Therefore, while one should not doubt that OGPL's advocated two- and threefold lexicographic process is indeed feasible, the way in which it is described, using just a CQS and a DWS, has not yet been implemented as such.⁸ Once it has, it would be informative to see several case studies of words 'followed through' from corpus to (monolingual and bilingual) pre-dictionary database to final (monolingual and bilingual) dictionary articles.

8. A powerful achievement

The time is past when the lexicographer, using only her own judgement and knowledge of ... (Atkins ... 2008: 201)

... cricket would construe the expression in these terms (... and Rundell 2008: 297)

Despite the fact that there is some room for improvement, it should be clear from the analysis that this does not alter the fact that *The Oxford Guide to Practical Lexicography* is a most powerful lexicographic book-length achievement. Above all, it is witty throughout, as when Patrick Hanks is referred to as Johnson's 'twenty-first-century counterpart' (p. 274) or when the modifier that precedes Anna Wierzbicka is 'reliably contrarian' (p. 434). Rare are cases where science and art meet. Here is an instance!

Endnotes

1. It is not clear whether or not the listings are meant to be in alphabetical order. Some are (Chapters 2, 4, 7, 8 (with one error), 9, 10 (with one error), 12), one is not (Chapter 1), while for others the R section is not, while the F section is (Chapters 3, 5, 6, 11). Also, the alphabetical order used in the Reading sections is generally different from that in the Bibliography for family names that start with a particle (i.e. 'de' and 'van'). For consistency, 'eds.' should not be used on p. 257.
2. For the sake of consistency, all references should mention the family name only, so '~~D.A.~~ Cruse (1986, 2004)' on p. 282, '~~Deirdre~~ Wilson (forthcoming)' on p. 285, '~~George~~ Lakoff and ~~Mark~~ Johnson's seminal *Metaphors We Live By* (1980)' on p. 290, '~~Charles~~ Fillmore (2003: 268)' on p. 409, and '~~Janet~~ Whitcut (1988: 53)' on p. 449. Also, on p. 44 one of the dates is wrong: 'Martin and Al ~~1988~~' → 1990; what is not yet published is better referred to with 'forthcoming', on p. 96: 'Prinsloo ~~2008~~' and 'Rundell and Atkins ~~2008~~'; while, on p. 464, it is not clear who the author is of pages '14-24' and '93-96' in 'Cowie (1987a)', given the latter is an edited collection.
3. One cross-reference, on pp. 210-211, is to a source outside the textbook.
4. The details for Biber 1990, on p. 67 n17, have been forgotten, however.
5. Rather than the dichotomy 'active/passive' (p. 506), 'encoding/decoding' seem to be the preferred terms elsewhere in OGPL. The latter pair is covered in the Index.
6. One of the measurement and management tools built into TshwaneLex in this regard is for example the Ruler tool (cf., for a theoretical background, De Schryver 2005).
7. Here too, TshwaneLex offers dedicated tools to help reverse either side of a bilingual dictionary, and to ensure that the reversibility principle is adhered to (cf. e.g. De Schryver 2008: 286 n6).
8. Similarly, OGPL is also vague on whether one is dealing with one or two Style Guide(s) during compilation (cf. e.g. pp. 359 vs. 363). It's probably two.

References

Dictionaries

- Atkins, B.T.S. and P. Robert (Eds.). 1978. *Collins-Robert French-English, English-French Dictionary / Robert-Collins dictionnaire français-anglais, anglais-français*. Glasgow: Collins / Paris: Le Robert.
- CRFD-8 (2006) = Atkins, B.T.S et al. 2006⁸. *Collins-Robert French Dictionary*. Glasgow: HarperCollins Publishers.
- MED-1 (2002) = Rundell, M. 2002. *Macmillan English Dictionary for Advanced Learners*. Oxford: Bloomsbury Publishing Plc.
- MED-2 (2007) = Rundell, M. 2007². *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan Education.

Other sources

- Atkins, B.T.S. 2002. Then and Now: Competence and Performance in 35 Years of Lexicography. Braasch, A. and C. Povlsen (Eds.). 2002. *Proceedings of the Tenth EURALEX International Con-*

- gress, EURALEX 2002, Copenhagen, Denmark, August 13–17, 2002: 1-28. Copenhagen: Københavns Universitet, Center for Sprogteknologi.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- BNC. 2008. *British National Corpus* [online]. <http://www.natcorp.ox.ac.uk/>.
- CLAWS. 2008. *CLAWS Part-of-Speech Tagger for English* [online]. <http://ucrel.lancs.ac.uk/claws/>.
- CNN. 30 July 2007. *Larry King Live — Triple-Murder Victims Mourned* [online]. <http://transcripts.cnn.com/TRANSCRIPTS/0707/30/lkl.01.html>.
- COBUILD. 2008. *Collins Birmingham University International Language Database* [online]. <http://www.cobuild.collins.co.uk/>.
- De Schryver, G.-M.** 2002. Web for/as Corpus: A Perspective for the African Languages. *Nordic Journal of African Studies* 11(2): 266-282.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- 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.
- De Schryver, G.-M.** 2008. Why Does Africa Need Sinclair? *International Journal of Lexicography* 21(3): 267-291.
- Fillmore, C.J.** 2003. Double-Decker Definitions: The Role of Frames in Meaning Explanations. *Sign Language Studies* 3(3): 263-295.
- Google. 2008. Google Search Engine [online]. <http://google.com/>.
- Hanks, P.** 2001. The Probable and the Possible: Lexicography in the Age of the Internet. Lee, S. (Ed.). 2001. *ASIALEX 2001 Proceedings, Asian Bilingualism and the Dictionary*: 1-15. Seoul: Yonsei University, Center for Linguistic Informatics Development.
- Hanks, P.** 2004. The Syntagmatics of Metaphor and Idiom. *International Journal of Lexicography* 17(3): 245-274.
- Hanks, P.** 2008. Lexical Patterns: From Hornby to Hunston and Beyond. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 89-129. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Hoey, M. and M.B. O'Donnell.** 2008. Lexicography, Grammar, and Textual Position. *International Journal of Lexicography* 21(3): 293-309.
- Ilsou, R.F. (Ed.).** 1986. *Lexicography: An Emerging International Profession*. Fulbright Papers 1. Manchester: Manchester University Press.
- Industry News*. 17 June 2002. Dragon Brands Defends the 'Consignia' Re-Brand [online]. http://www.brandintelligence.com/live/articles/dragon_defends_the_consignia_rebrand.asp.
- Johnson, S.** 1747. *The Plan of an English Dictionary*. Edited by Jack Lynch [online]. <http://andromeda.rutgers.edu/~jlynch/Texts/plan.html>.
- Johnson, S.** 1755. *Johnson, Preface to the Dictionary*. Edited by Jack Lynch [online]. <http://andromeda.rutgers.edu/~jlynch/Texts/preface.html>.
- Kilgarriff, A.** 1996. *BNC Database and Word Frequency Lists* [online]. <http://www.kilgarriff.co.uk/bnc-readme.html>.
- Kilgarriff, A.** 1997. Putting Frequencies in the Dictionary. *International Journal of Lexicography* 10(2): 135-155.

- Lewandowska-Tomaszczyk, B. and P. Hanks.** 1996. Completive Particles and Verbs of Closing in English. Weigand, E. and F. Hundsnurscher (Eds.). 1996. *Lexical Structures and Language Use, Volume 2*: 89-103. Tübingen: Niemeyer.
- LEXICOM. 2008 *Lexicography MasterClass — Lexicom Workshops* [online]. <http://www.lexmasterclass.com/>.
- Sinclair, J.M.** 1987. The Nature of the Evidence. Sinclair, J.M. (Ed.). 1987. *Looking Up, An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 150-159. London: Collins ELT.
- Sinclair, J.M.** 1998. The Lexical Item. Weigand, E. (Ed.). 1998. *Contrastive Lexical Semantics*: 1-24. Current Issues in Linguistic Theory 171. Amsterdam: John Benjamins.
- TshwaneLex*. 2008. TshwaneLex Dictionary Compilation and Terminology Management Suite [online]. <http://tshwanedje.com/tshwanelex/>.
- Urdang, L.** 1966. The Systems Designs and Devices Used to Process the Random House Dictionary of the English Language. *Computers and the Humanities* 1: 31-33.
- Urdang, L.** 1984. A Lexicographer's Adventures in Computing. *Dictionaries: Journal of the Dictionary Society of North America* 6: 150-165.
- Wikipedia*. 2008. Wikipedia — The Free Encyclopedia [online]. <http://www.wikipedia.org/>.
- Wiktionary*. 2008. Wiktionary — The Free Dictionary [online]. <http://www.wiktionary.org/>.
- WordSmith*. 2008. WordSmith Tools — Lexical Analysis Software for the PC [online]. <http://www.lexically.net/wordsmith/>.

Addendum 1: 'Hard words' in *The Oxford Guide to Practical Lexicography* (with Page = page number in OGPL; Freq., Word, POS and Files = frequency, word (i.e. type), part of speech, and number of files the word occurs in — all in the hundred-million-word BNC, as tagged by CLAWS (cf. Kilgarriff 1996), this to illustrate their infrequent use; this list obviously does not include words from citations, dictionary quotations or corpus lines)

Page	Freq.	Word	POS	Files	Page	Freq.	Word	POS	Files
43	325	amenable	aj0	245	179	2	siphoned	aj0-vvn	2
52	11	tetchily	av0	10		2	siphoned	vvd	2
68	120	creditable	aj0	102		2	siphoned	vvd-vvn	2
69, 434	37	egregious	aj0	32		20	siphoned	vvn	19
76	107	conscientiously	av0	96	285	77	amelioration	nn1	49
91	6	aggregating	aj0-vvg	4		1	amelioration	nn1-np0	1
	2	aggregating	nn1-vvg	2		1	ameliorations	nn2	1
	20	aggregating	vvg	18	285	—	pejoration		
95	47	unimpeachable	aj0	41	320	17	pre-ordained	aj0	17
157	—	anisomorphic				1	pre-ordained	aj0-nn1	1
166	60	chaff	nn1	44	323	57	multifarious	aj0	52
	2	chaff	nn1-np0	2	369	36	prescient	aj0	32
	12	chaff	nn1-vvb	10	426	75	epithet	nn1	64
168	8	phatic	aj0	7		3	epithet	nn1-vvb	3
168	69	nigh	av0	67	430	208	epistemological	aj0	54
	1	nigh	unc	1	437	2	ergative	aj0	2
	1	nigh/	unc	1	452	158	stricture	nn1	26
						1	stricture	nn1-vvb	1
						1	stricture(s)	nn2	1

Addendum 2: Errors in *The Oxford Guide to Practical Lexicography*

A. General

A.1 Typos

- 'a bilingual English–French dictionary' → a **unidirectional** bilingual ... (p. 24 n4)
- 'a bilingual English–French dictionary' → a **bidirectional** bilingual ... (p. 24 n5)
- **unshaded** → shaded (p. 209)
- Mel'čuk → Mel'čuk (p. 257)
- 'no hits at all **on** the BNC' → ... in the BNC (p. 412)
- 'So don't **to** try to force ...' → ... don't try to ... (p. 416)
- 'Meijs, W. ...' → Meijs, W. ... (p. 524)

A.2 Inflection and syntax

- '*adjudicate* sometimes **take** a direct object' → *adjudicate* sometimes takes ... (p. 60)
- 'Template entries are 'pro forma' entries for use by the lexicographers writing either database or dictionary.' → ... writing either **a** database or **a** dictionary (p. 123, plus similar cases)
- 'Apresjan (1973) described the semantic components which **gave** rise to ...' → ... give rise to ... (p. 139)

- 'rumours *circulate*, storms *rage*, heart *beats*' → ... a heart *beats* (p. 152)
- 'in this section we discuss the **factor** that influence these choices' → ... the factors that influence ... (p. 431)

A.3 Capitalization

- **EURALEX** is written in caps, all other -LEXes are not (p. 10)
- 'be used as **Linguistic data**' → ... linguistics data (p. 51)
- Word Sketch(es) ↔ **w**ord **s**ketch(es) (eg. pp. 107 *fn* vs. 369-370)
- 'those who treat dictionary users as if they are **L**exicographers' → ... as if they are lexicographers (p. 436)
- **A**bbreviation → abbreviation (p. 531)

A.4 Punctuation

- '(senses 1, 3, 4, and **5**).**This** technique ...' → ... and 5). This ... (p. 38)
- 'lexical-profiling software, for example (discussed in §4.3.1.5) only' → ...**,** only (p. 61)
- **etc** vs. etc. (e.g. pp. 104 vs. 105)
- '... and Moon 1998,**,** Mel'čuk 1998 presents ...' → ... Moon 1998. Mel'čuk 1998 ...' (p. 167)
-). → .) (pp. 182, 451)
- .) →). (p. 280)
- .) → .) (p. 461)
- 'including a Kiswahili-English one: (<http://africanlanguages.com/swahili/>)' → ... Kiswahili-English one (<http://africanlanguages.com/swahili/>) (p. 257)
- 'two broad semantic groupings: 'relationships' and documents' (or ...' → **,** documents' (p. 302)
- 'Part-specific PP-specific'. → last ' should not be in italics (p. 331)
- 'given in Box 9.#,**,** together ...' → ... in Box 9.#, together ... (pp. 331, 343, 347)

A.5 Inconsistencies, fonts and layout

- 'wording of definitions' vs. 'wording of the definitions' (e.g. pp. 35 vs. 38)
- bullets that do not match the ensuing text (e.g. pp. 57 *fn*)
- inclusion vs. exclusion of 'www' in URLs (e.g. pp. 83 vs. 96)
- fonts of figures need to be streamlined (e.g. Fig. 4.6, p. 108 vs. Fig. 4.10, p. 125)
- space needed between first and second group in Fig. 5.10 (p. 137)
- text needs to be broken up in extra paragraphs (pp. 215, 303, 373, 417, 461)
- 'collocate type' [sic] and 'collocate-type' is mentioned twice, in two different bullets (p. 322)
- meaning line should be dropped in Fig. 9.2 (p. 326)
- **I** *just stood ...* → 'I should be in italics (p. 335)
- 'one reason **for** **Mr Major's reluctance** to ...' → 'for' should be in bold (p. 341)
- construction 9 from Fig. 9.13 not in Box 9.3 (pp. 341-342 vs. 339-340)
- indentations for most quoted material are missing in Chapter 10 (pp. 393 *fn*)
- not used when pages face each other, so may be dropped: '(cont.)' (p. 432)

B. Specific

B.1 In quoted dictionary text

- missing space after headword (Fig. 6.14)
- missing spaces after tilde (Figg. 2.3, 7.6, 7.14)

- no space required after tilde (Fig. 12.7)
- missing space after label (Fig. 7.37)
- missing space after plus sign (Fig. 7.29)
- missing space after colon (Fig. 12.3)
- missing dot after POS (Fig. 7.16)
- missing dot in ellipsis (Fig. 7.28)
- bits should not be bold (Figg. 2.3, 6.2, 6.17, 6.19, 12.6)
- bit should be in bold (Fig. 12.10)
- wrong font (Fig. 12.10, 12.20)
- wrong colours (Figg. 7.3, 12.5, 12.22)
- missing colour (Fig. 12.7)
- misplaced colour (Fig. 7.17, 12.21)
- dot should be comma (Fig. 12.5)
- typo in 'nous retrouverons une solution' → nous trouverons ... (Figg. 2.7 and 2.8)
- missing space in 'ask sth(for sth)' → ask sth (for sth) (Fig. 5.24)
- missing words (Figg. 7.17, 7.28)
- star marks should be on the same level (Fig. 7.6)
- OALD examples from Fig. 7.28 do not come from the fifth edition
- reference to the dictionary itself inconsistent (e.g. Figg. 5.6 vs. 6.14)

B.2 In concordances

- .. → . (Fig. 6.5)
- missing dots (Fig. 6.7)
- in general: the use of full stops is not consistent (e.g. p. 301)

B.3 In graphs

- the graph does not correspond with the text (and truth) for the stretch 1960s–1970s (p. 58)

B.4 In boxes

- use of '+' vs. 'with' inconsistent (Boxes 9.2, 9.3, 9.4)
- use of 'infinitive verb' vs. just 'verb' inconsistent (Boxes 9.2, 9.3, 9.4)
- alphabetical sorting not fully correct (Boxes 9.3, 9.4)

Addendum 3: Suggestions for improvement in *The Oxford Guide to Practical Lexicography*

A. More conventional formulations

- 'the languages of southern Africa' → the Bantu languages (p. 2)
- 'superior number' → superscript number (p. 192)
- 'dedicated side columns' → Extra Column (p. 221)

B. Likely better

- Addition to the Reading section of Chapter 6: Kilgarriff 1997 (p. 199)
- It would have been a good idea to show concordances for 'happier' and 'happiest' as well in Fig. 9.15 (p. 345)

-
- In §9.2.10 one could have added that cross-references are also used between synonyms ('=') and antonyms ('><'), and to show words are related to one another ('<' and '>') (p. 378)
 - 'The nouns shown in square brackets' → ... in italics (p. 470)
- C. May be improved upon**
- The reader may wonder: What happened to the lemmatization phase? (p. 90)
 - The concepts 'homography', 'homophony' and 'homonymy' are not treated satisfactorily. (cf. esp. pp. 191-193, 280-282)
 - Excessive prudence is shown in the treatment of 'offensive terms': in contrast to all other terms that need labelling, not a single example is shown. (p. 229)
 - With reference to electronic dictionaries, OGPL claims: 'The field is developing so quickly that what we write today is almost guaranteed to be out of date by the time this book appears' (p. 238). This is a huge overstatement, as a comparison with even a five-year-old study of the field (De Schryver 2003) will easily show.
 - Box 9.1: Where is the field for frequency? Also note that nothing whatsoever is said about the manipulation of frequencies in the database, even though statements about rare vs. frequent occurrences abound. (p. 321)
 - With reference to a defining vocabulary (DV), one important issue is not mentioned, namely that expressed in number of senses, one *is* in fact dealing with most of the lexicon. (p. 449)
 - While §§11.2.1 and 11.2.2 are arguably brilliant, §11.3 disappoints in its simplicity. (p. 468 *fn*)
- D. Incorrect statements**
- With reference to Fig. 7.28, one reads: 'The note at the head of the *OHFD into* entry gives the English-speaking user some general advice about how to put this preposition in French' (pp. 234-235). This is incorrect. What the 'local usage note' does is to give guidance on where in the dictionary users need to search for the translation equivalents of the various constructions with *into*.
 - 'The homograph number is manually entered' → surely the software (a DWS?) automates this?! (p. 325)
 - In the 'Introduction to Part III' it is said that a methodology was described in Chapters 8 and 9 to build a database (correct) and to translate that database (incorrect). The latter is actually the topic of Chapter 11 only. (p. 383)
- E. Additional (recent) references (for a second edition)**
- To complement the discussion of discourse structure and rhetorical features (p. 81): Hoey and O'Donnell 2008
 - To complement the discussion of the coding systems lexicographers use(d) to indicate grammatical behaviour (p. 401, as well as Box 9.2 on pp. 332-333): Hanks 2008

Anna Dziemianko. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English.* 2006, XII + 229 pp. ISBN-13: 978-3-484-39130-7. ISBN-10: 3-484-39130-8. Lexicographica. Series Maior 130. Tübingen: Max Niemeyer. Price: €124.

To be user-friendly, monolingual and bilingual dictionaries for foreign users must be both easy to use and easy to read. They are designed to help prospective users both encode and decode, and to facilitate the comprehension of some vocabulary items or particular grammatical constructions. However, it often happens in practice that dictionaries are not used fully because some users are not aware of the richness of their contents. In carrying out the redesign of their dictionaries, lexicographers have been helped greatly by the views, needs, and preferences of a wide range of language users, many of them, of course, from schools and universities as well as general language learners. Further adaptations always follow, based on the experience of using a new dictionary and for this purpose feedback is always of great importance. In the process of establishing a user-friendly dictionary, the changes, though extensive, remain modest but in the much longer term, more complex changes take place to incorporate revisions and give them new clarity and coherence across the many and expanding contexts in which they are used. Whether this redesign actually helps users is the subject of much research. Both traditional and electronic dictionaries have now included a large number of tools in the definition, such as IPA pronunciation, examples and syntactic information, often presented in the shape of codes which will enable users not only to understand a particular structure, but to be able to reuse it.

Anna Dziemianko, the author of *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*, is Assistant Professor at the Instytut Filologii Angielskiej UAM in Poznan (Poland). This rigorous and meticulously documented monograph provides extensive research on the way modern, currently available English dictionaries have conveyed the syntactic behaviour of verbs and collocations, and have tried to ascertain whether their use of formal and functional-formal code labels is actually effective. There is a large gap in this area of research on verb syntax. Dziemianko questions the benefits of often oversophisticated and heavy code systems, especially for language users with limited dictionary skills. Statistical research has shown how cryptic the syntactic patterns presented by dictionaries can appear for the common user. The lack of standardisation between dictionaries is a further issue. After a trend that consisted in cumulating alphabetical or numerical combinations to such an extent that the definition might sometimes have six consecutive mnemonic codes which would have lost even the most dedicated students, the trend has been for dictionaries to attain a certain degree of simplification as well as uniformity between dictionaries. Full-sentence definitions in verb entries have now been incorporated to facilitate students' understanding of syntactic patterns and, although this will mean the lexicographer can dispense from using codes, certain information such as the transitiveness or intransitiveness of the

actual verb could prove to be not precise enough or be partly lost. With the advent of language corpora and masses of data extracted from the media, literature and the Internet, among other sources, contemporary monolingual and bilingual English dictionaries have sought a complete picture of the syntactic behaviour of verbs rather than an accurate description.

As the basis of her experiment, Dziemianko focuses on a selection of fifteen verbs, some more complex than others (like 'to anoint', 'to yank', and 'to subpoena') and identifies a certain number of variables (analytical versus contextual definitions, use of formal versus functional codes and localization of codes whether they are incorporated in the verb entry or stand out in an extra column) to measure the usability and user-friendliness of modern language learner's dictionaries[†] and gauge language users' strategies. Informants were asked to complete a multiple-choice test and to underline the part of the information they found useful to complete the exercise. The experiment was conducted on two groups of 300 students: high school Polish students and EFL students from Poznan University. Chapter 3 highlights the main findings of the experiments. Students with a higher proficiency standard tend to use multiple components of the entry, while the less proficient high school group does not. The higher level of proficiency in English undoubtedly makes the subjects' reading of syntactic patterns easier. The examples in the verb entry constituted for both groups the favoured piece of information, to the detriment of the actual definitions which were deemed to be the least useful. The physical location of codes did not seem to make any difference at all to the number of times they were consulted. Indeed, codes were consulted very often, though mostly by university students. The preference was shown to be for functional codes. On the whole, pattern illustrations in entries were frequently used, most particularly by university students. However, it transpires that their consultation could distract dictionary users from their main focus. The author brings to the fore interesting gender differences with female less advanced users resorting overall to codes more frequently than men, but more research would need to be conducted in order to corroborate these results. Dziemianko notes that 'a user-friendly verb entry should contain examples, a contextual definition and functional codes, interspersed with examples' (p. 188). The question, however, remains as to how to create a complete and accurate account of verb behaviour (considering some verbs having a whole range of patterns), without overloading the content of the verb entry and making it both opaque and abstract. A choice will have to be made between minimal information for the sake of clarity and transparency, and the more complex information for highly-proficient students. As she says, quoting Barone (1978: 188), 'even the most perfect dictionary is bound to remain a partial and inadequate instrument' (p. 190). In addition, there are still some points of grammar that dictionaries overlook and there is still much to be done to arrive at a complete and intelligible description. Perhaps electronic dictionaries permanently redesigned by lexicographers will allow, in their microstructure of entry, clarity and ease of use, and will not

require considerable familiarity with grammatical terminology and elaborate skills and training.

Unfortunately, for a book on user-friendliness this volume is not always easy to read. There is no index or glossary to guide the reader. Furthermore, the reader easily becomes lost amongst the use of acronyms, complex figures and tables interspersed throughout the book.

On the whole, Dziemianko provides some very valuable data and this is undoubtedly an important contribution to help us understand students' lookup behaviour and their success or failure in retrieving the information. Making the pedagogical dictionary more explicit will certainly help both language learners and their teachers.

Endnote

- † These include all editions of the *Oxford Advanced Learner's Dictionary of Current English* (1948, 1963, 1974, 1989, 1995, 2000) and of the *Longman Dictionary of Contemporary English* (1978, 1987, 1995, 2003), the three editions of the dictionaries published by Collins and based on the Collins Birmingham University International Language Database (COBUILD), the *Collins COBUILD English Language Dictionary* (1987), the *Collins COBUILD English Dictionary* (1995) and the *Collins COBUILD English Dictionary for Advanced Learners*, the two editions of the Cambridge dictionaries, the *Cambridge International Dictionary of English* (1995) and the *Cambridge Advanced Learners' Dictionary* (2003), and lastly the *Macmillan English Dictionary for Advanced Learners* (2002).

Reference

- Barone, Rosangela. 1978. On the Use of the *Advanced Learner's Dictionary of Current English*: From Theory to Practice. *Rassegna Italiana di Linguistica Applicata* 11: 187-194.

Michaël Abecassis
Department of Modern Languages
University of Oxford, Oxford
United Kingdom
(michael.abecassis@mod-langs.oxford.ac.uk)

Henrik Gottlieb and Jens Erik Mogensen (Editors). *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004.* 2007, XI + 321 pp. Terminology and Lexicography Research and Practice. Volume 10. ISBN 978-90-272-2334-0. Amsterdam/Philadelphia: John Benjamins. Price: €105.

This volume contains a selection of papers read at the 12th International Symposium on Lexicography, held in Copenhagen from April 29 to May 1, 2004. Many topical issues are discussed, and even though the volume is arranged according to different lexicographical themes, there are also many common elements between the different papers. For example, sample articles selected for the empirical analyses on which the papers are based, often deal with the use of phraseology in different dictionaries, the finding of suitable translation equivalents in bilingual dictionaries, and the user-friendliness in dictionaries. In some cases, the participants have selected the same mainstream dictionaries for discussion, which means that these dictionaries are examined from different viewpoints, making the volume an interesting and useful reference work.

The six parts into which the articles in this collection have been classified are online lexicography, dictionary structure, phraseology in dictionaries, LSP lexicography, dictionaries and the user, and etymology, history and culture in lexicography. The issues raised are of great importance to lexicographers, especially those working in dictionary houses and lexicography units where comprehensive lexicographical works are published and compiled, such as multi-volume corpus-based monolingual dictionaries and comprehensive bilingual dictionaries. It is taken for granted that technology is part of the lexicographical process nowadays, and therefore the papers in this volume are most relevant.

The main speaker was Arne Zettersten who is also the 'founding father' of this symposium (see p. ix). His paper 'Glimpses of the Future of English-based Lexicography' (pp. 299-318), found in the last part of the volume, is the Otto Jespersen Memorial Lecture, dedicated to Otto Jespersen for his direct and indirect role in promoting lexicography. Zettersten contemplates on the future use of English in lexicography, and the nature of future lexicographical works. Since English is currently a lingua franca in many spheres of the modern world, including technology and the internet, Zettersten proceeds from the assumption that this will still be the case, even when other languages such as Chinese Mandarin, Spanish, Arabic and Hindi will gain in importance. Therefore, the question 'What kind of English or Englishes do we expect in the world tomorrow?' will have to be considered. He reflects on how quickly current electronic sources such as the Encarta products by Microsoft (e.g. the *Encarta World English Dictionary Online*) and *The Literary Encyclopedia* (released in 2001) will become outdated, and what possibilities the future might hold. Especially in the field of technology, the growth of vocabulary is accelerating all the time, and printed reference works cannot keep up with the pace of change. In addition, new technical devices are produced very quickly, which opens up exciting

possibilities for lexicography. For example, the contingency for users to have direct access to a corpus, or to update a personalised dictionary automatically, together with the rapid globalisation and the need to communicate easily and efficiently across linguistic and cultural borders, ensures a bright future for digital lexicography. In fact, since this paper was presented, another four years of technological advances have passed — apart from the advances in the technology of mobile phones since then, other devices are currently available for storing and using books and reference works. For example, one such a device, the Reader Digital Book by Sony (<http://www.sonystyle.com>), can be used as the 'perfect travel companion', since it is compact and lightweight, and has a battery life for reading 7 500 pages. It can hold about 160 eBooks, and has the possibility of optional memory cards. Electronic books, and by implication, electronic dictionaries, could be downloaded, and then read on the 6 inch display, which has a paper-like texture. The device, which makes searching and browsing possible, is compatible with formats such as Adobe PDF and Microsoft products. These types of technologies of course have far-reaching implications for the planning and publication of lexicographical works.

The second plenary lecture, also found in the last part of the volume, was given by Henning Bergenholtz and Sven Tarp, under the title 'Politik und Sprachpolitik in der Lexikographie' (= 'Politics and Language Politics in Lexicography') (pp. 217-240). They distinguish between (1) language politics in general, which include for example the promotion of a particular language, that may sometimes be conservative or purist, having the goal of promoting one as official language against other competing languages in the same geographical area, (2) specific language politics which deal with communication strategies and stylistic applications, in order that specific communication may take place, for example political correctness and the extent to which lexicographers use descriptive and/or prescriptive approaches, and (3) conscious and fixed selection or recommendation of specific grammatical constructions, words or word forms in a particular language and its implications for lexicographical activities such as corpus planning. This last-mentioned type of language politics undoubtedly is very important for lexicographers. Bergenholtz and Tarp (p. 219) claim that even a 'descriptive' approach can be conservative, and, by implication, misleading for users, referring to the 'prescriptive power of descriptive statements' (p. 238, note 1). Decisions taken in lexicography have language-political relevance, and therefore a political dimension. To illustrate this point, they analyse examples of Danish expressions and admit that in different languages, different language-political customs prevail.

The third plenary paper, found in the fifth part of the volume, was read by Herbert Ernst Wiegand. His contribution is entitled 'Neuere Aspekte einer Theorie und Typologie von Wörterbuchartikeln und ihre Praxisrelevanz' (= 'Newer Aspects of a Theory and Typology of Dictionary Articles and their Practical Relevance') (pp. 183-200). Wiegand takes an in-depth look at the term *Wörterbuchartikel* (= *dictionary article*), revisiting his own definition which was

given more than 20 years ago (Wiegand 1984). He does this because he discovered that there was no genus proximum such as the English hyperonym *basic reference unit* under which this German term can be accommodated. Over the years, Wiegand has refined his theory of dictionary structures to such an extent that he can now distinguish much more subtly and point out a much larger number of constituents in dictionary articles. After having investigated 1 200 dictionaries over the years, he introduces terminology for a semantic network into which the term *Wörterbuchartikel* can be assigned a place. By means of tree graphs, he makes distinctions to illustrate this terminology via typology criteria, such as classes of types of reference works, types of outer access structures, the nature of accessibility, the number of possible access paths, the nature of the access text element, the availability of a system of arranging entries, the nature of such a system, the presence or lack of standardised text condensation, the presence of item texts, the degree of text condensation, and the presentation of the propositional article centre. Wiegand concludes that a *dictionary article* is a lexicographical accessible entry which has the following three features: (1) it exhibits at least one access text element; (2) it is a constituent of the word list; and (3) it consists of accessible lexicographical data, by means of which the dictionary user can open up lexicographical information mentioned by the access text element about the subject matter of the reference work.

The first section of the volume is devoted to the topic of online lexicography. Cristina Gelpí looks at the 'Reliability of Online Bilingual Dictionaries' (pp. 1-12). Dictionary quality is determined by factors such as the authors of the online dictionary and the quality of their work, the satisfaction of the communicative needs of the users, i.e. the usability of the dictionary, the frequency of updating the dictionary, and the degree of digitisation and hypertextuality. Not all online dictionaries have been conceptualised as online products, with the result that they do not use the full potential of digital resources, such as interactiveness or incorporating image, sound and video material. The paper by Annette Klosa and Carolin Müller-Spitzer, 'Grammatische Angaben in *lexiko* und ihre Modellierung' (= 'Grammatical Items in *lexiko* and their Modelling') (pp. 13-37) deals with a specific dictionary project at the Institut für Deutsche Sprache (IDS) in Mannheim. This project aims at compiling an online database for the conducting of online research. This database which will be regularly updated and expanded, consists mainly of text excerpts collected from newspapers and magazines, representing text types such as travel reports, readers' letters, advertisements, extracts from novels, official texts, etc. The *lexiko* database is part of the IDS corpus, the world's largest corpus for the German language, having almost two billion text words. *Lexiko* is planned in such a way that it can offer the possibility to indicate frequency, as in corpus research. Not only will the part of speech of a lemma be given, but also its syntactical relationships, valency features or morphological forms. The advantages over traditional non-computerised lexicography are clear. By using the corpus, 'synchronic "empirical" grammatical research' (p. 20) can really be done

because such up-to-date language use is documented in it. One of the disadvantages is that if the corpus has gaps, *ellexiko* will also show these gaps. The modelling of *ellexiko* entails that the contents and layout have to be planned separately, as is usually the case with online publications. The authors illustrate the types of decisions to be made in the modelling of this database by means of examples. For instance, it is important that the data can be presented in different ways, and for different situations of use, but draw on the same lexicographical database. It is also important to take into account the different types of research possibilities which users may expect. Włodzimierz Sobkowiak's paper 'Innovative Phonetic Interfaces for Electronic Dictionaries' (pp. 39-52) also deals with technology-driven lexicography. His investigation concerns the possibilities that innovative visualisation techniques may have for the presentation of pronunciation in learners' electronic dictionaries of English as a foreign language (EFL). With new techniques such as animation, colour, hyperlinks, 3D views, and drag-and-drops, Sobkowiak argues for presentations of the pronunciation e.g. by means of a visualisation he calls a *graphic frame*. This graphic frame represents a two-dimensional 'neighbourhood slice' of a larger and more complex knowledge structure. For example, the pronunciation of *said* can be linked to the pronunciations of other concepts such as *sad, say, head, bed, Port Said, laid*, etc., in order to help foreign language learners distinguish between them, while showing the entire network of concepts, with some written comments and guidelines, on the computer screen, and presenting the possibility for sound clips for each of these concepts.

The second section of the volume deals with aspects of dictionary structure. Rufus Gouws's paper on 'Sublemmata or Main Lemmata: A Critical Look at the Presentation of Some Macrostructural Elements' (pp. 55-69) discusses some space-saving procedures in Afrikaans monolingual and English-Afrikaans bilingual dictionary articles which at times seem inconsistent because of the nesting. Consequently it is not always clear whether a particular simplex or compound is a main lemma or a sublemma. User needs are not always taken into account, and therefore in the case of idioms, users cannot easily determine whether they have to do with an idiom or whether it is a usage example. In bilingual dictionary articles, there is often no indication whether an idiom in the source language also has an equivalent idiom in the target language, and whether the style and register is the same in both languages. Gouws argues that this practice stems from a linguistic approach as opposed to a more pragmatic approach which takes the users' needs into account. The paper of Andrejs Veisbergs is entitled 'Semantic Aspects of Reversal of a Set of Bilingual Dictionaries' (pp. 71-79), dealing with the process of reversal of translation equivalents to enable lexicographers of a new Latvian-English dictionary to ascertain whether they have taken all the possible equivalents into account. Although dictionary reversal is a controversial practice, the results of this particular project described by Veisbergs did produce an increase in material. Naturally, some serious editing had to be done, and it had to be checked whether each

item was used frequently enough to justify its inclusion in the dictionary. Some of the problems encountered when reversing such a bilingual dictionary were: (1) explanatory translations presented for some language items where no translation equivalents are available become invalid; (2) differences in style and connotation, e.g. in the case of slang words, have to be considered; (3) even denotations can differ where 'semantically broad entries' have 'narrower, more concrete counterparts' (p. 77).

In the third section of the volume, phraseology in dictionaries is discussed. Ken Farø's paper treats 'Idiomatische Äquivalenzprobleme: Ein ikonoklastischer Zugang' (= 'Problems of Idiomatic Equivalence: An Iconoclastic Approach') (pp. 83-95). The problem of phraseological equivalence is also dealt with by Erla Hallsteinsdóttir in her paper 'A Bilingual Electronic Dictionary of Idioms' (pp. 97-106), which furthermore looks at the presentation of idioms. It seems that idioms still pose many problems in bilingual lexicography, and Hallsteinsdóttir investigates the potential and advantages of computerised approaches.

Section four of the volume focuses on LSP lexicography. The first paper by Patrick Leroyer is entitled 'Bringing Corporate Dictionary Design into Accord with Corporate Image' (pp. 110-117). According to him, corporate communication has specific needs which can only be satisfied by planning the messages they send out and documenting these messages in a dictionary for their specific corporate use. Communication data can be classified according to themes, and the in-house dictionaries should be able to assist users in finding the correct terms when documents in different genres have to be drawn up. It should also assist them in multilingual situations, e.g. by localising the message for the particular context. Staying within the corporate world, Sandro Nielsen and Lise Mourier deal with the 'Design of a Function-based Internet Accounting Dictionary' (pp. 119-135). The accounting dictionary under discussion is conceptualised for Danish users, and therefore the descriptive information is in Danish. The data types for instance include definitions and factual information about the lemma, and recommended usage, grammatical information, markers for varieties of English (such as British or American), equivalents, cross-references, synonyms and antonyms, and sources (including internet links), Danish collocations and phrases with English translations, etc. The authors claim that this unique dictionary meets the requirements of a state-of-the-art electronic bilingual internet LSP dictionary, and illustrate this by means of examples.

Dictionaries and the user is the topic of the fifth section of the volume. Here, Ilan J Kernerman asks the question: 'What's So Good or Bad about Advanced EFL Dictionaries?' (pp. 139-145). Kernerman proceeds from the point of view that today there are more non-native than native speakers of English — and this is growing steadily. These EFL speakers are not foreigners, but 'people around the world with their own languages, using English as part of their rapidly changing reality, and changing the English language therewith' (p. 140), making EFL a communication tool in the global network of commerce and

other activities. He looks at some examples from well-known EFL dictionaries such as the *Longman Dictionary of Contemporary English* (LCODE, Summers (Ed.) 1995), and poses the question whether they still fulfil the real needs of their users. The following paper by Renata Szczepaniak, entitled 'Glimpses into Receptive Dictionary Use: Proficient Learners' Perspective' (pp. 147-164), also uses examples from the LCODE, but with a view to highly advanced learners of English. She too deals with idioms, and specifically the problems which can occur when the 'canonical' meaning of idioms are changed in texts (p. 150). One such example is 'tarring the whole of the Austrian public with the Haider brush is unfair'. Szczepaniak tries to determine how this type of change in meaning affects EFL readers' receptive abilities. In the last paper in this section, Geert van der Meer discusses 'The Learners' Dictionaries and Grammar' (pp. 165-181). He also analyzes dictionary articles from five major English learner's dictionaries, as regards the specific problems occurring with grammatical information. Native speakers of English would not normally need this type of information, but for EFL users, it is very important. For example, determining the difference between *some* and *any* is not always easy because EFL speakers do not know whether they denote an indefinite number or a definite amount. The so-called *he/she* problem is another example, as well as when and how to use *which* and *that* in relative clauses. Van der Meer analyses a number of these examples, and in each case gives some suggestions together with a sample dictionary article which could solve the problems.

Part 6 of the volume deals with etymology, history and culture in lexicography. Arleta Adamska-Salaciak's paper 'Lexicographers as Borrowers — The Importance of Being CAMP²' (pp. 203-215) discusses the 'lack of interlingual equivalence' by taking the word *camp* and all its meanings in English as an example. The term as such shows an 'anisomorphism' in Polish, since its translation equivalents *kamp* or *kampowy* are mainly used in more academic types of texts, whereas this is not the case in English. According to the author, the equivalents in existing bilingual English-Polish dictionaries are very imprecise. Andreas Gröger's paper, 'Revising German Etymologies in the *Oxford English Dictionary*' (pp. 241-257), gives an account of the procedures involved in the revision of the etymological information for the third edition of the *Oxford English Dictionary*. Lexicographers can learn much about the painstaking thoroughness of following up references and sources described here, and the discoveries resulting from this. 'Extracting Usage Information from 18th Century English-Danish and Danish-English Dictionaries' is the title of the paper by Marcin Overgaard Ptaszyński (pp. 259-276). He discusses several English-Danish dictionaries from the eighteenth century in which the first labels were used. Inconsistencies, misprints and mistranslations are pointed out, and he claims that some of the problems, such as labelling the difference between regional words and regional concepts, are still present today. Roda P. Roberts's paper entitled 'Dictionaries and Culture' (pp. 277-297) deals with two monolingual dictionaries of English, *Random House Webster's Unabridged* (1999) and *The*

Oxford Canadian (Barber (Ed.) 2001), and two bilingual English–French dictionaries, *Robert-Collins Senior* (Atkins et al. 1996) and *Oxford-Hachette* (Corréard and Grundy (Eds.) 2001). The items with geographical labels pertaining to national culture, superculture and subculture are analysed, as well as definitions, encyclopedic notes, and linguistic usage notes.

All in all, this volume makes interesting and useful reading for lexicographers working at different projects. It is certainly informative to discover what other lexicographers are engaged in, and how they solve their particular problems. Common elements, especially within the framework of computer-assisted and computerised lexicographical projects, can be found throughout the volume.

References

- Atkins, B. et al. (Eds.). 1996. *Le Robert and Collins Dictionnaire français–anglais, anglais–français Senior*. New York: HarperCollins/Paris: Dictionnaires Le Robert.
- Barber, K. (Ed.). 2001. *The Canadian Oxford Dictionary*. Toronto: Oxford University Press.
- Corréard, M.-H. and V. Grundy (Eds.). 2001. *The Oxford-Hachette French Dictionary*. Oxford: Oxford University Press.
- Random House Webster's Unabridged Dictionary*. 1999. New York: Random House.
- Summers, D. (Ed.). 1995. *Longman Dictionary of Contemporary English*. Third Edition. Harlow: Longman.
- Wiegand, H.E. 1984. Prinzipien und Methoden historischer Lexikographie. Besch, W., O. Reichmann and S. Sonderegger (Eds.). 1984. *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*: 557-620. Handbuch zur Sprach- und Kommunikationswissenschaft 2.1. Berlin/New York: De Gruyter.

Maria Smit
Music Education
Department of Music
University of Stellenbosch
Stellenbosch
Republic of South Africa
(msmit@sun.ac.za)

Paul Newman. *A Hausa–English Dictionary*. 2007, x + 243 pp. ISBN 978-0-300-12246-6. New Haven/London: Yale University Press. Price: £40.

Hausa (Chadic/Afroasiatic) is a major world language, spoken by more than 40 million people who are mostly found in northern Nigeria and the Republic of Niger. A sizeable number reside in other parts of Nigeria and the major cities of West Africa (Accra, Kumasi, Douala, Cotonou) and beyond. The history of its documentation dates back more than 150 years, with the most comprehensive grammars of any African language (see Wolff 1993, Jaggard 2001, for example). In the context of dictionary making, the language has also been the subject of lexicographic interest for more than a century, ranging from the pioneering work of Robinson (1899) to the large and accurate dictionary by Bargery (1934). Hausa has also had the benefit of more than a dozen other dictionaries, including Abraham (1962), Newman and Newman (1977), Mijinguini (1987), Newman (1990), Caron and Amfani (1997), Awde (1996) and more recently CNHN (2006). All these efforts have been acknowledged by the author in his introductory remarks. And it is against this comprehensive treatment of Hausa lexicography that the proper place of Newman's *A Hausa–English Dictionary* needs to be examined.

The aim of *A Hausa–English Dictionary* as spelt out in the introductory remarks (p. ix) is to serve as a counterpart to R.M. Newman's excellent *An English–Hausa Dictionary*, with the target group being 'language learners and practical users', particularly those with 'familiarity and grasp of the fundamentals of Hausa grammar'. In the practical user group, it must be assumed, are not only 'Hausa-language scholars and students', as the blurb points out, but also native Hausa speakers, particularly speakers of non-standard varieties. It can even be argued that anyone interested in African language documentation and description will benefit from the concise grammatical sketch provided in the introduction (pp. ix–xxii) as well as in the appendices (pp. 237–243). These 'notes', it should be pointed out, are the outcome of a lifelong passion for and study of Hausa language and linguistics, which had earlier seen the appearance of *The Hausa Language: An Encyclopedic Reference Grammar* (Newman 2000).

A Hausa–English Dictionary is based on a highly systematized lexical database of some 12 000 head entries, encompassing not only 'new ... terminology ... of the modern world', but also a unique and unparalleled interest in the contextual use of these entries in idiomatic, figurative and specialized usage. I will say more on these later. Although the major Hausa-speaking assistants who participated in the project came from the major Hausa dialect areas (Kano, Sokoto and Niger), the entries themselves are overwhelmingly based on (standard) Kano Hausa, the written variety encountered as every-day Hausa. Taken together, these elements help to make this work an outstanding contribution to Hausa scholarship, even by Paul Newman's exacting standards.

In addition to its excellent typeface, with its bold head entries and column-based presentation, a number of useful innovations are noticed. The introduc-

tion consists of more than ten pages of detailed explanation of the nature and description of the dictionary entries. These are not just stylistic issues related to alphabetization and the transcription system (p. x), but also substantial information concerning the nature of the language. Thus, plurality (p. xiv, par. 5.1.3) is dealt with not only as a specific (dictionary) issue, such as how plural nouns are coded (e.g. *jàakii* 'donkey', m. (f. -aa) <-una>), but also as an aspect of the language system, analyzing how Hausa plurals and their varying forms are formed as a step-by-step process. The same detailed and analytical description is also evident from such Hausa peculiarities as verbal nouns and the grade system. Thus, even in these introductory pages, Newman exhibits his 'facts count' formula on how to practise linguistics as a science (see Newman 1991, Newman 2000a and Newman 2007).

Another very useful indicator is the introduction of additional lexicographic labels to accompany the standard abbreviations and symbols table. Thus, the labels are employed to highlight context of use, cultural nuances and appropriateness. Labels such as erudite (e.g. *alfyàn* the number '2 000'), polite (e.g. *nàjasàa* 'excrement'), pejorative (e.g. *ìnyaamuri* 'Igbo') and vulgar (e.g. *gatò* 'vagina') are not ordinarily stressed in African language lexicography, and this must be taken as signalling the coming of age of dictionary writing in these languages. Hausa society, with its stress on 'kunyàa' (embarrassment, shame, modesty, propriety) in public discourse, employs a large vocabulary of euphemisms, often Arabic-based, to speak about sex and related bodily functions, so it is useful that these are documented (and there are many more in this dictionary!), but labelled to indicate the degree of sensitivity towards their usage. (Curiously, the most recent monolingual *Kamusun Hausa* (CNHN 2006) dictionary from the Centre for the Study of Nigerian Languages [CSNL] exhibits no such objections even to some of the most taboo words in Hausa, e.g. *tseelee* 'cunt' [not in my Word 2004 for Mac English dictionary and thesaurus, nor in *A Hausa-English Dictionary*!].)

Similarly, dialectal variants, such as *hànwaaawàa* = *hàwainìyaa* 'chameleon', *jìidà* = *sàuka* 'descend, come down' and alternative usages, such as '*hakkii* = *hakkii* 'right, entitlement', *bàree* = *bàllee* 'let alone', *fùròofeesàa* = *fàrfeesàa* 'professor', and *wàatò* = *wàatàu* 'in other words, that is to say', are indicated, since the fairly homogenous nature of Hausa means that there is a fine line to be drawn between standard (written) Hausa and its dialectal variants in the core Hausa-speaking north-west of Nigeria (but see further comments below). One of the ways in which these variations manifest themselves in everyday Hausa is in the metaphorical extensions of core meanings, an area ably treated in *A Hausa-English Dictionary*. Two examples will show how Newman handles these descriptions.

The verb CI is defined as 'eat' (pp. 33-35), followed by its standard exemplification in *wutaa taa ci rùmfaa* 'fire consumed the shed'. Then follows a long list of more than fourteen derivative (cognitive/figurative/idiomatic) meanings, such as (abrade) *igiyàa taa ci kafar àkuyàa* 'the rope burned the goat's leg';

(win, overcome) *ci nasaràa* 'be victorious', *ci jarràbâtúwaa* 'pass an examination'; (enjoy, experience) *ci kàasuúwaa* 'have a successful day at the market', *ci duuniyàa* 'enjoy life'; (idiom) *ci àmaanàa* 'breach trust'. Related compound forms are also entered independently, as they occur, e.g. *cí-baaya* 'regression', *cí-gàba* 'progress', *cín-hancù* 'bribery', *cù-raani* 'migrant labour' (see Jaggar and Buba 2007).

Similarly, the highly productive *dan* diminutive modifier has over sixty independent entries (pp. 50-52). The entries describe its various adjectival functions, including familiar usages such as *dan-adàm* 'human' (cf. *haKKìn-dan-adàm* 'human rights' (p. 84)), *dan-giyàa* 'drunkard', also its journalistic appropriation, as in *dan-gudùn-hijiraa* 'refugee', *dan-taakaraa* 'candidate', *dan-tà-kifèe* 'risk-taker', as well as the relatively more recent (post-R.M. Newman 1990) *dàn-àcaadàa* 'motorcycle taxi driver', *dàn-à-moorèe* 'vandal' (but not the much older *dàn-à-waarèe* 'secessionist' usage) (see McIntyre and Meyer-Bahlburg 1991).

These entries are a measure of the breadth and depth of *A Hausa-English Dictionary*, making it unassailably 'the dictionary of choice for Hausa-language scholars', as William R. Leben of Stanford University correctly underscores in his comment on the back cover. They also show the high level of accuracy in Newman's tone and vowel length transcription, hence underscoring the essential role of tone and vowel length in extracting pragmatic information from Hausa dictionary entries.

Now it is a fact that these detailed descriptions may not be useful or even necessary for all languages. It is possible, for example, that for the less well-documented languages of the world, basic lexical entries with a short and concise meaning are enough, but for Hausa, with its over 40 million first and second language speakers and excellent grammars and dictionaries, this dictionary has succeeded in achieving the right balance between those learners interested in basic definitions and the academic and teaching communities whose focus is much more technical. In a recent web posting, a seasoned Hausaist sought help on how to translate 'choices/options/elections' into Hausa, since Newman has (correctly) pointed out in his dictionary that *zàaɓɛe* is commonly used generically. It is for the benefit of professionals such as these that a detailed and extended explanation is desirable. Extended meanings and erudite usage, as exemplified in *A Hausa-English Dictionary*, will ensure that a greater democratization of knowledge continues to be paramount in the minds of lexicographers and (language) policy makers. It is instructive, however, that a number of words labelled as erudite by the compiler turned out to be unfamiliar to many of my bilingual university linguistics undergraduates (see below). It may be imagined that instead of using these words in speaking about and describing technical information, students will resort to English equivalents leading to the common phenomenon of Hausa-language code-mixing, popularly known as *Inghausa* i.e. English-Hausa (see Madaki 1983).

In this respect, the compiler deserves commendation for his large inventory of modern and modernized Hausa words currently making their impact within modern Hausa society. Localized terms such as *bùrookàa* 'Pidgin Eng-

lish', *goo-sùlò* 'traffic jam', *màjoooratèe* '[political] majority', *ògaa* 'boss, chief executive officer', *cíf-joojì* 'chief judge of a high court', *galàn* 'pot hole; gallon', not to forget local terms such as *Bùhàariyyàa* 'austerity period' (during Gen. Buhari's military regime in the 1980s), *'yar-Gùsau* '1940s famine, feeling hungry', *gidan-gàba* 'front seat of car' and *gidan-baaya* 'back seat of car' (but not *tà-zarcèe* 'self-succession at all cost') as well as transitional words such as *eejèn* 'agent', *têf* 'tape recorder, measurement tape', *daftàn* 'Dr.' (= Ph.D.), *naasàrèe* 'nursery', *tíi* 'tea', and *ii-mèl* 'e-mail' (but not *intàneeti* 'internet', *tùurís* 'imported second hand car', *ànkòo* 'uniformed wedding party dressing', or *tiibìi* 'TV') have been documented, as are the extended meanings of common terms, such as *danyee* 'inter/fresh', *dàn-kooree* 'hustler', and *dàn-baarikii* 'city slicker'.

Ironically, aspects of its strength may turn out to be a weakness in other respects. Erudite terminology forms a large class of the vocabulary in this dictionary. This is essential for anyone seeking to localize the specialized education, health, policy and communications register. As I remarked earlier, this helps to avoid the choice of unassimilated English loanwords, some of which may never become localized, e.g. 'mobile' (but see *hansìt* 'handset' and *sàluulaa* 'cellular'). The problem with this class of specialized terms, however, is that a number of native speakers do not seem to be familiar with them, let alone use them. The following is an example from a list of randomly extracted vocabulary (in A–C), that neither I nor my final year (Hausa) linguistics students recognized initially: *àlàtuurèe* 'protest', *àlayyàdii* 'oil from palm kernel', *àlgàràagîs* 'type of small ring-shaped sweet bread', *àlhiinii* 'sad thoughts, meditation', *akumàarii* 'pad for pack animal', *àzàlzalàa* 'pestering', *azarfaabii* 'overeagerness', *bàdoodàraa* 'slovenly woman with big buttocks', *baagarwaa* 'mere show; matter of unimportance', *càssaa* 'bowleggedness', *ceedâa* 'vomiting'. Nor are some of these items either in the most recent monolingual *Kamusun Hausa* (CNHN 2006) (e.g. *àlàtuurèe*, *àzàlzalàa*), or in Bergery's classic *Hausa–English Dictionary* (1934) (e.g. *azarfaabii*, *bàdoodàraa*). Many of these obscure forms are of Arabic origin, and have been a major source of (specialized) Hausa loanwords. Thus, it is not so much a matter that words in this category are undocumented, but that they are no longer in current usage, perhaps because of their association with an older material culture. It could also be caused by the fact that much of the unfamiliar terminology originate from the dictionary's heavy reliance on a 'standard' Kano dialect database. I say 'standard', because the number of these unfamiliar terms in *A Hausa–English Dictionary* leads me to the conclusion that the colloquial *Kananci* dialect continues to be (erroneously?) equated with standard (written) Hausa in dictionaries and grammars. Note however that the author has acknowledged the Kano connection of the database from the outset. Nor should this observation invalidate the comprehensive array of dialectal (*d.v.*), alternative and colloquial variants in this dictionary.

During the eight months of an intensive use of *A Hausa–English Dictionary* in an advanced Hausa class, there have only been a few occasions where my students and I have failed to find forms we came across during the course of

our classes (see below). Even vocatives such as *hêe!* 'hey you!, pay attention!' *hoohòo!* 'what a pity!, what bad luck!', *kayya!* 'expresses doubt', and female-specifics, such as *àhâyyee!* 'expression of pleasant joking between women', *ayyuru-rui!* 'the sound of shrilling ... esp. at a wedding', *lâalee!* 'welcome!', and *bâmbu* 'invitation or command to a baby for it to be carried on the mother's back' have been presented. On many occasions, my students and I were impressed by the large number of *Sakkwatançi* (and other dialect forms) that we have found in the dictionary, and perhaps more importantly for me, some have no dialectal (*d.v.*) labelling (see *bùkwii* 'bare', for example). Also, typographical errors are so few that only the most painstaking detective work could uncover **attakaa* (*àtak-kàa* 'energy'), **bàbambađee* (*bàbambaađee* 'professional cadger'), **islaamà* (*islaamà* 'islamic') and **bil-hakkì* (*bil-hakkì* 'seriously, in earnest').

On the plus side, too, it needs to be noted that the relatively recent (re)implementation of the *Shari'a* system in many parts of the core Hausa-speaking northern Nigeria is increasingly helping to popularize and reintroduce this classical Arabic-based terminology. This in turn has had a beneficial impact on the growth of mainly Hausa internet forums, such as *Finafinan Hausa* (Hausa Films Discussion Forum) and *Marubuta* (Hausa Writings Forum) where debates about the impact of *Shari'a* on Hausa popular culture continue to boost the range of classical vocabulary available to participants. Thus, Newman's *A Hausa-English Dictionary* may be anticipating a new era of significant expansion of the language, not so much from the usual (English) source, but from the dormant (and dominant) Islamic discourses of today's northern Nigeria.

The other source of this expansion must be the explosive growth in new Hausa writing, commonly termed *soyayya* (love) novels, which have engendered increasing literacy, especially among monolingual Hausa women, who constitute both the majority of the writers as well as the readers of this genre (see Furniss 1996, Furniss, Buba and Burgess 2004).

There then remain a small number of words whose figurative meanings have not been documented. The verb *caaɓɓa* 'make slushy or splotchy', for example, is more commonly used for 'blunders' than the literal meaning given. (See also *hurùmii* 'land ... with restricted use' (to the exclusion of its judicial usage of 'court jurisdiction').) Similarly, commonly used words such as reduplicated *gùrgùje* 'hurried', *daakìlee* 'dimmed', *maslahàa* 'truce', *gudàawaa* 'diarrhea' and *tàatacce* 'infamous, notorious' are rare omissions in this otherwise comprehensive work. In other cases, the meanings of tonally-distinguished minimal pairs are confused, as in *halii* 'character', *haalii* 'opportunity, chance', and *haalii* 'means', although in almost all other respects *A Hausa-English Dictionary* is an excellent source of tone discrimination for Hausa students and teachers alike, using the large inventory of such minimal pairs (see *baba*, for example, with the following inventory: *bàaba* 'father', *baabà* 'mother', *baabàa* 'barber', *baabaa* 'indigo tree' and *bàabaa* 'eunuch!').

In other words, nothing in this review should detract from the commendable presentation in *A Hausa-English Dictionary*. Its accurate tone and vowel-

length transcriptions, its large modern, colloquial and ideophonic inventory, coupled with an acute sensitivity to the appropriate meanings of words would surely make it an example of how dictionaries should be compiled. As the Hausa saying goes: *yàbon gwàni yaa zama doolè* 'praising the expert is a duty'. I have no hesitation in recommending *A Hausa–English Dictionary* by Paul Newman (Malam Sabo (lit. 'New Man'!) as he is popularly known in the Hausa world).

References

- Abraham, Roy Clive.** 1962. *Dictionary of the Hausa Language*. Second Edition. London: University of London Press.
- Awde, Nicholas.** 1996. *Hausa–English English–Hausa Dictionary*. New York: Hippocrene Books.
- Bargery, G.P.** 1934. *A Hausa–English Dictionary and English–Hausa Vocabulary*. London: Oxford University Press
- Caron, Bernard and Ahmed H. Amfani.** 1997. *Dictionnaire français–haoussa suivi d'un index haoussa–français*. Paris: Karthala.
- CNHN.** 2006. *Kamusun Hausa*. Kano: Cibiyar Nazarin Harsunan Nijeriya, Jami'ar Bayero.
- Furniss, Graham.** 1996. *Poetry, Prose and Popular Culture in Hausa*. Edinburgh: Edinburgh University Press.
- Furniss, Graham, Malami Buba and William Burgess.** 2004. *Bibliography of Hausa Popular Fiction*. Cologne: Rüdiger Köppe.
- Jaggar, Philip J.** 2001. *Hausa*. Amsterdam/Philadelphia: John Benjamins.
- Jaggar, Philip J. and Malami Buba.** 2007. Metaphorical Extensions of 'eat' ⇒ [overcome] and 'drink' ⇒ [undergo] in Hausa. Newman, John (Ed.). *The Linguistics of Eating and Drinking*. Amsterdam/Philadelphia: John Benjamins.
- Madaki, Rufa'i Omar.** 1983. *A Linguistic and Pragmatic Analysis of Hausa-English Code-switching*. Ph.D. Thesis. Michigan: University of Michigan.
- Mijinguini, Abdou.** 1987. *Karamin Kamus na Hausa zuwa Faransanci*. Niamey: CELHTO.
- McIntyre, Joseph and Hilke Meyer-Bahlburg, assisted by Ahmed Tijani Lawal.** 1991. *Hausa in the Media: A Lexical Guide*. Hamburg: Helmut Buske.
- Newman, Paul and Roxana Ma Newman.** 1977. *Modern Hausa–English Dictionary*. Ibadan/Zaria: Oxford University Press (Nigeria).
- Newman, Paul.** 1991. Facts Count: An Empiricist Looks at Indirect Objects in Hausa. Hubbard, Kathleen (Ed.). 1991. *Berkeley Linguistic Society: Proceedings of the Annual Meeting. Special Session on African Language Structures, February 15–18, 1991*: 155-165. Berkeley: University of California, Berkeley Linguistics Society.
- Newman, Paul.** 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven: Yale University Press.
- Newman, Paul.** 2000a. Writing a Reference Grammar of an African Language: Conceptual and Methodological Issues. Wolff, H. Ekkehard and Orin D. Gensler (Eds.). 2000. *Proceedings of the Second World Congress of African Linguistics. Leipzig, 1997*: 33-47. Cologne: Rüdiger Köppe.
- Newman, Paul.** 2007. [With Alan S. Kaye.] An Interview with Paul Newman. *Semiotica* 166: 237-278.

Newman, Roxana Ma. 1990. *An English–Hausa Dictionary*. Yale: Yale University Press.

Robinson, C.H. 1899. *Dictionary of the Hausa Language*. Cambridge: Cambridge University Press.

Wolff, H. Ekkehard. 1993. *Referenzgrammatik des Hausa*. Hamburger Beiträge zur Afrikanistik 2. Münster/Hamburg: Lit.

Malami Buba
Department of Modern European Languages and Linguistics
Usmanu Danfodiyo University
Sokoto
Nigeria
(mb4383@aol.com)

Publikasieaankondigings / Publication Announcements

- B.T. Sue Atkins and Michael Rundell. *The Oxford Guide to Practical Lexicography*. 2008, 552 pp. ISBN-10: 0199277710. ISBN-13: 9780199277711. Oxford: Oxford University Press. Price: £29.99. (Review article in this issue.)
- Lynne Bowker (Editor). *Lexicography, Terminology and Translation: Text-based Studies in Honour of Ingrid Meyer*. 2006, xii + 252 pp. ISBN-13: 978-0-7766-0627-9. ISBN-10: 0-7766-0627-1. Ottawa: University of Ottawa Press. Price: \$35.
- Gilles-Maurice de Schryver (Editor-in-chief). *Pukuntšu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. 2007, xii + 552 + R1-16 pp. ISBN 978 0 19 576555 7. Cape Town: Oxford University Press South Africa. Price: R99.95.
- Madaleine du Plessis (Hoofredakteur/Chief Editor). *Pharos 1, Afrikaans–Engels/English–Afrikaans Woordeboek/Dictionary*. Eerste uitgawe/First edition. 2005, xx + 1552 pp. ISBN 1868900444. Kaapstad/Cape Town: Pharos Woordeboeke/Pharos Dictionaries. Prys/Price: R450. (See article in this issue.)
- Anna Dziemianko. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. 2006, XII + 229 pp. ISBN-13: 978-3-484-39130-7. ISBN-10: 3-484-39130-8. Lexicographica. Series Maior 130. Tübingen: Max Niemeyer. Price: €124. (Review in this issue.)
- Thierry Fontenelle (Editor). *Practical Lexicography: A Reader*. 2008, 405 pp. ISBN 978-0-19-929234-9. Oxford Linguistics. Oxford: Oxford University Press. Price: £27,50.
- Henrik Gottlieb and Jens Erik Mogensen (Editors). *Dictionary Visions, Research and Practice. Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004*. 2007, XI + 321 pp. Terminology and Lexicography Research and Practice. Volume 10. ISBN 978 90 272 2334 0. Amsterdam/Philadelphia: John Benjamins. Price: €105 (Review in this issue.)
- Kusujiro Miyoshi. *Johnson's and Webster's Verbal Examples: With Special Reference to Exemplifying Usage in Dictionary Entries*. 2007, XIV + 222 pp. ISBN 978-3-484-39132-1. Tübingen: Max Niemeyer. Price €82.

- Marijke Mooijaart and Marijke van der Wal (Editors). *Yesterday's Words: Contemporary, Current and Future Lexicography*. 2008, viii + 376 pp. ISBN (10): 1-84718-469-3. ISBN (13): 9781847184696. Newcastle: Cambridge Scholars Publishing. Price: £39.99.
- Paul Newman. *A Hausa–English Dictionary*. 2007, x + 243 pp. ISBN 978-0-300-12246-6. New Haven/London: Yale University Press. Price: £40. (Review in this issue.)
- Fred Pheiffer (Redakteur/Editor). *Pharos Afrikaans–Engels/English–Afrikaans Kernwoordeboek/Concise Dictionary*. Eerste uitgawe/First edition. 2007, xx + 1620 pp. ISBN-10: 1-86890-048-7. ISBN-13: 978-1-86890-048-0. Kaapstad/Cape Town: Pharos Woordboeke/Pharos Dictionaries. Prys/Price: R375. (See article in this issue.)
- Pumlani Sibula. *Isigama soMthetho/Law Terminology/Regsterminologie*. 2007, x + 56 pp. ISBN 978-0-620-40270-5. Stellenbosch: IZiko leLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.
- A.J. Smit (Hoofredakteur), W.A. Landman, A.E. Gerber en P.L. de Jager (Redaksionele Komitee). *Opvoedkundewoordeboek. Vertalend, Verklarend. Afrikaans/Engels. Deel 2: H–O*. 2007, ii + 440 pp. *Deel 3: P–Z*. 2007, iii + 313 pp. ISBN 0 620 33683 8. Selfuitgawe: A.J. Smit, Trevorstraat 33, Murrayfield 0184.
- Sven Tarp. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. 2008, VIII + 308 pp. ISBN 978-3-484-39134-5. Lexicographica. Series Maior 134. Tübingen: Max Niemeyer. Price: €84.
- Herbert Ernst Wiegand. *Internationale Bibliographie zur germanistischen Lexikographie und Wörterbuchforschung. S–Z. Mit Berücksichtigung anglistischer, nordistischer, romanistischer, slavistischer und weiterer metalexikographischer Forschungen. Band 3: S–Z*. 2007, VIII + 508 pp. ISBN 978-3-11-019027-4. Berlin/New York: Walter de Gruyter. Price: €228.

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met die Buro van die WAT in verbinding (wat@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.sun.ac.za/wat>)

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 rekenaarskyf (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 the Bureau of the WAT (wat@sun.ac.za) or refer to our website: <http://www.sun.ac.za/wat>)

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 typewritten 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 mit dem Büro des WAT Kontakt auf (wat@sun.ac.za) für eine ausführlichere Wiedergabe dieser Hinweise oder besuchen Sie unsere Webseite: <http://www.sun.ac.za/wat>)

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 selbstständig, 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 (wat@sun.ac.za)
ou consulter notre website: <http://www.sun.ac.za/wat>)

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*.