

Lexikos 17

Lexikos 17

Redakteur

Editor

J.C.M.D. du Plessis

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 17:2007

AFRILEX SERIES 17:2007



BURO VAN DIE WAT

STELLENBOSCH

Die publikasie van hierdie boek is moontlik gemaak deur 'n ruim subsidie van die L.W. Hiemstra-Trust.

The publication of this book was made possible by a generous subsidy from the L.W. Hiemstra Trust.

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2007 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2007

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: 978-0-9584971-5-2
ISBN-10: 0-9584971-5-X
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, 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)
Prof. E.F. Kotzé (RSA)
Dr. P.A. Louw (RSA)
Prof. L.J. Louwrens (RSA)
Mnr. K.J. Mashamaite (RSA)
Dr. P.A. Mavoungou (Gaboen/Gabon)
Dr. J.S. Mdee (Tanzanië/Tanzania)
Prof. B.M. Mini (RSA)
Mnr. M.C. Mphahlele (RSA)
Mnr. M.H. Mpungose (RSA)
Prof. D.J. Prinsloo (RSA)
Prof. P.H. Swanepoel (RSA)

Inhoud / Contents

Voorwoord	xi
Foreword	xiii
<i>J.C.M.D. du Plessis</i>	
'n Woord van AFRILEX	xv
A Few Words from AFRILEX	xvi
Redaksionele doelstellings	xvii
Editorial Objectives	xviii
Redaktionelle Ziele	xix
Politique éditoriale	xx

Artikels / Articles

Log Files Can and Should be Prepared for a Functionalistic Approach	1
<i>Henning Bergenholtz and Mia Johnsen</i>	
Professionele vertalers as woordeboekgebruikers	21
<i>M.F. Crafford en Ilse Feinauer</i>	
Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho	37
<i>Gilles-Maurice de Schryver and Elsabé Taljard</i>	
Terminologiebestuur in Suid-Afrika. 'n Beoordeling van drie werkmodes	56
<i>Dina Ferreira en Theodorus du Plessis</i>	
A Transtextual Approach to Lexicographic Functions	77
<i>Rufus H. Gouwus</i>	
Étude socio-terminologique du vocabulaire médical cilubà	88
<i>Emmanuel Kambaja Musampa</i>	
Kiswahili Verbs: A Lexicographical Challenge	114
<i>Z.s.m. Mochiwa</i>	

E-dictionaries and Phonolexicographic Needs of EFL Users <i>Włodzimierz Sobkowiak</i>	134
Issues in the Planning and Design of a Bilingual (English– Northern Sotho) Explanatory Dictionary for Industrial Electronics <i>Elsabé Taljard, Rachéle Gauton and Liam A. Gauton</i>	152
Lexicography in the Information Age <i>Sven Tarp</i>	170
Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform <i>Herbert Ernst Wiegand</i>	180

Lexikoprogrammatuur / Lexicosoftware

Creating a Single South African Keyboard Layout to Promote Language <i>Dwayne Bailey</i>	212
Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of <i>TshwaneLex</i> <i>Gilles-Maurice de Schryver and Guy De Pauw</i>	226

Resensieartikels / Review Articles

Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs <i>Michaël Abecassis</i>	247
Doen die <i>Woordeboek van die Afrikaanse Taal Deel XII</i> dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie? <i>Ilse Feinauer</i>	259

Leksikofokus / Lexicofocus

The African Languages Research Institute: A Milestone in the Development of the Zimbabwean Languages <i>Emmanuel Chabata</i>	278
--	-----

Language Planning and Monolingual Dictionaries: With Special Reference to Ndebele <i>Samukele Hadebe</i>	292
Considerations for Providing Etymological Information in the <i>Advanced Ndebele Dictionary</i> <i>Langa Khumalo</i>	307
Lexicographical Practice and Lexicological Research: The Case of Shangani in Zimbabwe <i>Peniah Mabaso</i>	316
Diglossic Principles in Shona Lexicography <i>Esau Mangoya</i>	329
The Role of the African Languages Research Institute in Addressing Language of Instruction Dilemmas in Zimbabwe <i>Jesta Masuku and Finex Ndhlovu</i>	340
<i>Isichazamazwi SesiNdebele</i> as Reflector of the Moral and Ideological Values of Society <i>Nobuhle Moyo</i>	349
The User Perspective in Lexicography: The Lemmatisation of Fixed Expressions in <i>Duramazwi Guru reChiShona</i> <i>Nomalanga Mpofu</i>	360
Some Deficiencies in the Mediostructure of <i>Isichazamazwi SesiNdebele</i> <i>Dion Nkomo</i>	372
The Impact of Lexicographical Work on Language Use: The Case of Shona Monolingual Dictionaries in Zimbabwe <i>S. Nyota and J. Mapara</i>	385
Projekte / Projects	
Furthering the Aim of Multilingualism through Integrated Terminology Development <i>Pumlani M. Sibula</i>	397

Leksikonotas / Lexiconotes

- On the Timelessness of Music Dictionaries 407
Henning Bergenholtz and Inger Bergenholtz
- Implementing a Dictionary Culture in South Africa: An Attempt
at a General Framework for the African Languages 416
Juliane Klein
- Kinship Terminology: Problems in Some English–Tshivenda
Bilingual Dictionaries 424
Munzhedzi James Mafela
- Polysemy and Homonymy: Challenges Relating to Lexical
Entries in the *Sesotho sa Leboa–English Bilingual Dictionary* 433
V.M. Mojela

Resensies / Reviews

- Albert Busch: *Diskurslexikologie und Sprachgeschichte der
Computertechnologie* 440
Andrzej Katny
- Julie Coleman: *A History of Cant and Slang Dictionaries. Volume I:
1567–1785* 442
Julie Coleman: *A History of Cant and Slang Dictionaries. Volume II:
1785–1858*
Anne McDermott
- Christa Kilian-Hatz: *Khwe Dictionary* 446
Andy Chebanne
- Phillip Louw: *Oxford Afrikaans–Engels/English–Afrikaans
Skoolwoordeboek/School Dictionary* 453
Anton F. Prinsloo
- M.O. Mbatha: *Isichazamazwi SesiZulu* 458
Rose Masubelele
- Anton F. Prinsloo: *Sleng: Woorde, uitdrukkings en hul herkoms* 462
A.E. Cloete

Włodzimierz Sobkowiak: <i>Phonetics of EFL Dictionary Definitions</i> <i>Gilles-Maurice de Schryver</i>	466
Birgit Steinbügl: <i>Deutsch–englische Kollokationen: Erfassung in zweisprachigen Wörterbüchern und Grenzen der korpusbasierten Analyse</i> <i>Maria Smit</i>	471
H.C. Viljoen: <i>Kuberwoordeboek Afrikaans–Engels, Engels–Afrikaans / Cyber Dictionary Afrikaans–English, English–Afrikaans</i> <i>Liezl Potgieter</i>	478
Ferdie Weich: <i>San Dictionary/San-woordeboek. San–Afrikaans– English/English–San–Afrikaans/Afrikaans–San–English</i> <i>Andy Chebanne</i>	481
Herbert Ernst Wiegand: <i>Internationale Bibliographie zur germanis- tischen Lexikographie und Wörterbuchforschung</i> <i>Maria Smit</i>	487
Publikasieaankondigings / Publication Announcements	495
Voorskrifte aan Skrywers	499
Instructions to Authors	500
Richtlinien für Autoren	501
Instructions aux Auteurs	502

Voorwoord

Met die dood vanjaar van prof. Ladislav Zgusta het *Lexikos* een van sy mees vooraanstaande Adviesraadslede verloor. Sy *Manual of Lexicography* wat in 1971 verskyn het, was moontlik en is steeds een van die mees aangehaalde en vermelde publikasies in leksikografiese besprekings. Vir hom was woorde en woordeboeke belangrik, maar selfs belangriker was die mense agter die woorde en woordeboeke. Hy het 'n besondere belangstelling in die Suid-Afrikaanse leksikografie gehad. Gedurende April 1989 het hy die Universiteit van Stellenbosch besoek waar hy lesings gegee en ook aan 'n leksikografiese werksessie deelgeneem het. Daarna was daar voortdurende kontak omdat hy ingelig wou bly oor leksikografiese ontwikkelinge in Suid-Afrika. Die verlies van sy kundigheid sal 'n groot leemte laat.

Vanjaar het prof. Arthur Delbridge ook die wens uitgespreek om as lid van die Adviesraad van *Lexikos* uit te tree. Onder leksikograwe is prof. Delbridge veral bekend vir sy oprigting en voltooiing van die woordeboek van Australiese Engels, die *Macquarie Dictionary*. Na die bekendstelling van die derde uitgawe van die woordeboek in 1997, het hy uitgetree as Hoofredakteur, 'n pos wat hy vir byna 28 jaar beklee het. Sedert 1997 was prof. Delbridge 'n lid van die Adviesraad van *Lexikos*. *Lexikos* wil hom bedank vir die tien jaar wat hy deel was van die Adviesraad.

Prof. Fredric Dolezal van die V.S.A. en dr. Jan Tent van Australië het die uitnodiging aanvaar om lede van die Adviesraad te word. Prof. Arleta Adam-ska-Sałaciak van Pole het ingestem om Oos-Europa op die Adviesraad te verteenwoordig. *Lexikos* wil hierdie nuwe lede verwelkom en die wens uitspreek dat hulle 'n lang en vrugbare verbintenis met *Lexikos* sal hê.

In hierdie nommer word 'n nuwe afdeling "Leksikofokus" aangebied. Vanjaar is die fokus op die African Languages Research Institute (ALRI) in Harare, Zimbabwe. Hierdie Instituut het veral bekend geword vir sy leksikografiese werk soos blyk uit die verskeie woordeboeke wat alreeds deur sy navorsers gepubliseer is. In tien bydraes wat strek van langer artikels tot korter notas, gee die skrywers, meestal verbonde aan of geskakel met ALRI, nie net 'n oorsig van die aktiwiteite van ALRI, sy bevordering van die inheemste tale, sy rol in taalbeplanning, sy invloed in die onderwys en die trefkrag van sy publikasies op taalgebruik nie, maar bespreek ook sekere aspekte van of probleme betreffende sy gepubliseerde en beplande algemene woordeboeke soos etimologiese inligting, diglossiese beginsels, ideologiese kwessies, vaste uitdrukkings en mediostrukturele gebreke.

Behalwe hierdie bykomende afdeling, bevat hierdie nommer van *Lexikos* ook die afdelings wat gewoonlik deel vorm van sy inhoud, wat tot gevolg gehad het dat dit 'n uitgebreide en omvattende spesiale nommer geword het.

Voorwoord

Die redakteur wil al die bydraers en keurders bedank dat hulle hierdie nommer van *Lexikos* moontlik gemaak het. Beide bydraers en keurders het 'n belangrike aandeel aan die handhawing van die gehalte van *Lexikos*.

'n Spesiale woord van dank moet weer, soos in die verlede, aan me. Riette Ruthven gaan wat met haar bedrewe en professionele setwerk verantwoordelik was vir die aantreklike, maklik leesbare artikels en oorsigtelike tabelle. Me. Tanja Harteveld moet ook bedank word vir die oplossing van verskeie rekenaarmoelikhede.

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

Foreword

With the death this year of Prof. Ladislav Zgusta, *Lexikos* has lost one of its most distinguished Advisory Board members. His *Manual of Lexicography*, which appeared 1971, has probably been and still is one of the most quoted and referred to publications in lexicographical discussions. For him, words and dictionaries were important, but even more important were the people behind the words and dictionaries. He had a particular interest in South African lexicography. During April 1989, he visited the University of Stellenbosch where he lectured and also took part in a lexicographical workshop. After that there was continued contact for he wanted to remain informed about lexicographical developments in South Africa. The loss of his expertise will leave a big void.

This year Prof. Arthur Delbridge also expressed the wish to step down as a member of the Advisory Board of *Lexikos*. Among lexicographers Prof. Delbridge is noted particularly for his founding and completion of the dictionary of Australian English, the *Macquarie Dictionary*. After the launch of the third edition of the dictionary in 1997 he retired as Editor-in-Chief, a position he had filled for almost 28 years. Since 1997 Prof. Delbridge has been a member of the Advisory Board of *Lexikos*. *Lexikos* wants to thank him for the ten years he has been part of the Advisory Board.

Prof. Fredric Dolezal of the U.S.A. and Dr Jan Tent of Australia have accepted the invitation to become members of the Advisory Board. Prof Arleta Adamska-Salaciak from Poland has agreed to represent Eastern Europe on the Advisory Board. *Lexikos* wants to welcome these new members and to express the wish that they will have a long and fruitful association with *Lexikos*.

In this issue, a new section "Lexicofocus" is presented. This year the focus is on the African Languages Research Institute (ALRI) in Harare, Zimbabwe. This Institute has become known especially for its lexicographical work as shown by the several dictionaries already published by its researchers. In ten contributions, ranging from longer articles to shorter notes, the writers, mostly attached to or associated with ALRI, not only give an overview of the activities of ALRI, its promotion of indigenous languages, its role in language planning, its influence in education and the impact of its publications on language use, but also discuss certain aspects of and problems regarding its published and planned general dictionaries such as etymological information, diglossic principles, ideological issues, fixed expressions and mediostuctural deficiencies.

Apart from this additional section, *Lexikos* also presents the sections which usually form part of its contents, which resulted in it becoming an extensive and comprehensive special issue. The editor wants to thank all the contributors and referees for making this issue of *Lexikos* possible. Both contributors and

Foreword

referees have an important part in maintaining the quality and standard of *Lexikos*.

A special word of thanks must again, as in the past, go to Ms Riette Ruthven who, with her competent and professional typesetting, was responsible for the attractive, easily readable articles and conveniently arranged tables. Ms Tanja Harteveld must also be thanked for clearing up several computational difficulties.

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

'n Woord van AFRILEX

Dit is nogmaals 'n voorreg vir AFRILEX om met hierdie uitgawe van *Lexikos* vereenselwig te word — enersyds oor die gehalte van die teoretiese werk en die verslag oor die aard en omvang van die praktiese leksikografie wat in hierdie uitgawe aangebied word maar andersyds veral ook oor die bestek van die bydraes en bydraers. AFRILEX is 'n vereniging vir Afrika maar ook vir leksikograwe van buite Afrika. En dit blyk uit hierdie uitgawe. Dit is goed om 'n verskeidenheid bydraes van buite Afrika in te sluit maar dit is ook belangrik dat bydraes vanuit Afrika nie beperk is tot Suid-Afrika nie. Veral verblydend is die toevoeging van 'n nuwe afdeling "Leksikofokus" met die klem in hierdie uitgawe op die werk deur die ALRI-projek in Zimbabwe. Ons hartlike dank aan elkeen wat in hierdie uitgawe van *Lexikos* help om AFRILEX se doelwitte en doelstellings te bereik.

Vanjaar is dit tien jaar sedert dr. J.C.M.D. du Plessis die redakteurskap van *Lexikos* oorgeneem het en *Lexikos* 17 is die elfde uitgawe wat hy redigeer. In hierdie tien jaar het *Lexikos* sy plek en status as vaktydskrif op internasionale gebied bevestig en versterk. Vandag kan dit ongetwyfeld as een van die leidende vaktydskrifte vir leksikografie beskou word. Dit is in 'n hoë mate aan dr. Du Plessis se kundigheid, ywer, noukeurigheid en toewyding te danke. Namens AFRILEX wil ek aan hom ons opregte dank en waardering uitspreek vir die onbaatsugtige manier waarop hy hierdie taak verrig.

Ook AFRILEX betreur die dood van prof. Ladislav Zgusta. Vir almal van ons was hy 'n voorbeeld, vir baie van ons was hy 'n mentor en vir 'n paar bevoorregtes was hy 'n vriend. Ons bring hulde.

Die Dertiende Jaarlikse Internasionale Kongres van AFRILEX word beplan vir 30 Junie–3 Julie 2008. Die Buro van die Woordeboek van die Afrikaanse Taal in Stellenbosch tree as gasheer op. Meer inligting is beskikbaar op die AFRILEX tuisblad by <http://afrilex.africanlanguages.com>.

R.H. Gouws
President

A Few Words from AFRILEX

It is yet again a privilege for AFRILEX to be associated with this edition of *Lexikos* — on the one hand because of the quality of the theoretical work and the report on the nature and extent of practical lexicography on offer in this edition but, on the other hand, also the range of the contributions and contributors. AFRILEX is an association for Africa but also for lexicographers from outside Africa. And this is illustrated in this volume. It is good to include a variety of contributions from outside Africa but it is also important that the African contributions are not restricted to South Africa. Especially noteworthy in this regard is the introduction of a new section "Lexikofocus", with the emphasis in this issue on the work done at the ALRI project in Zimbabwe. Our sincere thanks to everyone who co-operated in this issue of *Lexikos* to help achieve the aims and objectives of AFRILEX.

This year it is ten years since Dr. J.C.M.D. du Plessis became editor of *Lexikos* and *Lexikos* 17 is the eleventh issue to be edited by him. During these ten years *Lexikos* has established and strengthened its position and status as scientific journal on the international level. Today it can undoubtedly be regarded as one of the leading journals for lexicography. This is to a large extent due to the expertise, diligence, preciseness and devotion of Dr. Du Plessis. On behalf of AFRILEX, I would like to express our sincere thanks and appreciation for the selfless way in which he performs this task.

AFRILEX also mourns the death of Prof. Ladislav Zgusta. He was an example to all of us, for many of us he was a mentor and for a few privileged ones he was a friend. We pay tribute.

The Thirteenth Annual International Conference of AFRILEX is planned for 30 June–3 July 2008. The Bureau of the Woordeboek van die Afrikaanse Taal in Stellenbosch will host this conference. More information is available on the AFRILEX website at <http://afrilex.africanlanguages.com>

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 uitgawe 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.

Log Files Can and Should Be Prepared for a Functionalistic Approach

Henning Bergenholtz, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa and Centre for Lexicography, Aarhus School of Business, Aarhus, Denmark (hb@asb.dk)*
and

Mia Johnsen, *Centre for Lexicography, Aarhus School of Business, Aarhus, Denmark (miajohnsen@gmail.com)*

Abstract: User surveys of printed dictionaries may be characterised as non-representative and non-realistic laboratory tests, often with retrospective questions based on memory. Log file analyses concerning the use of Internet dictionaries, on the other hand, are based on large numbers of users and look-ups. However, log file analyses have also been characterised by a juggling of numbers based on data calculations of limited direct relevance to practical and theoretical lexicography. This article proposes the development of lexicographically relevant log files for the use in log file analyses in order to give a true picture of how and why different dictionaries are employed for different purposes.

Keywords: LEXICOGRAPHY, LOG FILES, DICTIONARY, INTERNET DICTIONARY, SEARCH OPTIONS, DICTIONARY FUNCTION, RECEPTION, TEXT PRODUCTION, TRANSLATION, COMMUNICATIVE FUNCTION, COGNITIVE FUNCTION, DICTIONARY USE, USER SURVEY, LEMMA, DICTIONARY ITEM, USER NEEDS

Opsomming: Loglêers kan en behoort voorberei te word vir 'n funksionalistiese benadering. Gebruikersopnames van gedrukte woordeboeke kan gekarakteriseer word as nieverteenvoordigende en nierealistiese laboratoriumtoetse, dikwels met retrospektiewe vrae gebaseer op geheue. Loglêeronntledings betreffende die gebruik van Internetwoordeboeke, aan die ander kant, is gebaseer op groot getalle gebruikers en raadplegings. Loglêeronntledings word egter ook gekenmerk deur 'n gegoël met getalle gebaseer op databerekening van beperkte direkte tersaaklikheid vir die praktiese en teoretiese leksikografie. Hierdie artikel stel die ontwikkeling van leksikografies tersaaklike loglêers voor vir gebruik in loglêeronntledings om 'n ware beeld te gee van hoe en waarom verskillende woordeboeke vir verskillende doeleindes aangewend word.

Sleutelwoorde: LEKSIKOGRAFIE, LOGLÊERS, WOORDEBOEK, INTERNETWOORDEBOEK, SOEKKEUSES, WOORDEBOEFUNKSIE, ONTVANGS, TEKSPRODUKSIE, VERTALING, KOMMUNIKATIEWE FUNKSIE, KOGNITIEWE FUNKSIE, WOORDEBOEKGEBRUIK, GEBRUIKERSOPNAME, LEMMA, WOORDEBOEKITEM, GEBRUIKERSBEHOEFTE

1. Better Dictionaries through the Use of Log Files

The last decade has seen an explosive growth in the number of available online dictionaries. Significant technological improvements have led not only to an increase in the use of the Internet through high-speed connections, but also to new opportunities in the field of lexicography. In recent years, the focus has shifted from being mainly on the dictionary itself and the compilation thereof to the dictionary user — what does the user expect from the dictionary, and how do lexicographers best cater for the users' needs? This is one of the pivotal issues in the lexicographic debate and one that may be addressed through the use of the new technological possibilities.

The advent of online dictionaries has given rise to new ways of studying user habits and needs in order to improve a given dictionary in accordance with its users' wishes. Previously, lexicographers were forced to resort to such methods as user surveys, tests or protocols, which, due to their inherently subjective nature, do not necessarily provide a realistic picture. With online dictionaries, two new approaches have become accepted: log files and direct feedback. These methods may be used concurrently, and as De Schryver and Joffe (2004) describe in their article on the SeDiPro project, the results gained from the two methods will often supplement each other. In other instances, however, as was the case in direct feedback from users of the *Danish Internet Dictionary*, users request rarely used words which are not evident from the log files to be included in the dictionary, and the methods, rather than the results, may thus supplement each other.

One contribution on the use of log files for improving Internet dictionaries is Bergenholtz and Johnsen (2005). The article demonstrates that log files may be used to this end in various ways, most importantly to discover so-called lemma lacunae (i.e. words that could/should have been included in the dictionary), but also to discover other problematic issues. The aim is not to give a detailed account of this article, but merely to sum up the main points being made about this type of log file analysis.

As for lemma lacunae, the implications of log file analyses are obvious. Lexicographers can periodically analyse the log files and add words users have searched for without finding them, thus increasing the hit rate and the usability of the dictionary. This practice was also used in the SeDiPro project, resulting in an increase in the hit rate from 67% to 75% (De Schryver and Joffe 2004). Which words to include depends on the dictionary's functions, intended user group and genuine purpose and whether it is composed as a minimising or a maximising dictionary. Log file analyses may also reveal frequent searches for 'missing' orthographic forms. In Bergenholtz and Johnsen (2005), the passive and the imperative are mentioned as examples that may subsequently be added to the dictionary, thus making it possible for the users to search for these forms. This has, in fact, been done in the *Danish Internet Dictionary* as a result of log file analyses and feedback from users. Frequent misspellings or searches for

non-existing words are also revealed in the log files and provide lexicographers with various options in terms of helping the user. The incorrect word may be added to the dictionary with a reference to the correct term, or the user may automatically be redirected to the correct form of the word. These strategies are, of course, only feasible in the case of frequently occurring mistakes.

There is no doubt that log file analysis is an extremely useful tool in discovering and providing a solution to such issues as those described above. The greatest advantage of the method, however, its objectivity, also constitutes a limitation as the hard data do not reveal the motivation for a search, e.g. whether the user in question uses the dictionary in connection with reception, production or translation, or whether the user actually found the answer to his/her question. This issue will be discussed later in this article.

2. Log Files, Corpus-based Lemma Selection and Search Options

In a recent article by De Schryver et al. (2006), a Swahili–English dictionary project is described and the use of corpus-based lemma selection is criticised on a number of counts. The intention was to investigate whether users really look up the most frequent words in a corpus on the basis of an analysis of the log files for the dictionary, the authors concluding that this is not the case. The same conclusion was reached by Johnsen (2005) in an analysis of the log files for the *Danish Internet Dictionary*. This relation between corpus-based lexicography and log files is very interesting as it challenges a generally accepted and widely used approach to lemmatisation. In De Schryver et al. (2006), it is concluded that corpus-based lemma selection is a valid strategy in connection with minimising dictionaries, but not in the case of maximising dictionaries in which users also expect to find less common words. The authors suggest that corpora may be used as guidance, but that additional software modules should be applied to help users, e.g. a module for redirecting frequent misspellings as also suggested by Bergenholtz and Johnsen (2005) and a module that deals with multi-word units (MWUs). The use of this latter module does indeed seem to be a valuable tool, and the problem posed by MWUs gives rise to a discussion of the definition of a lemma and which units should be awarded lemma status in a given dictionary.

In our consideration, log files are a useful supplement to corpus-based lemma selection as they may be used to reveal lemma lacunae, frequent misspellings, frequent searches for MWUs, etc. Whether the shortcomings of corpus-based lemma selection should be remedied on the basis of log file analysis, through the use of software modules or a combination of these is for the lexicographers to decide, but it does indeed stand to reason to employ the new technological possibilities offered by the Internet. Spelling and typing mistakes may, for example, be rectified through an integrated automatic spell checker as in common text editing programmes such as Microsoft Word (cf. Bergenholtz 2005 on the status of Word's spell checker as a dictionary in its own right), or

the user may be presented with alternative suggestions if the search string is not found as is the case in the search engine Google. The latter strategy has also been implemented in the *Danish Internet Dictionary* where the user is presented with a list of 10 alternative suggestions ranked according to match frequency in the case of unsuccessful searches. The percentages shown after each suggestion indicate the probability of the suggestion being the correct alternative to the search string entered by the user. If the user searches for a word with incorrect spelling and the spelling mistake is a minor one with just one incorrect letter, the correct spelling of the word will usually occur as the first, and thus most likely, suggestion. This for example applies if the user searches for the incorrect spelling *hiraki* (hierarchy, the correct spelling being *hierarki*), in which case the following result appears (translations by HB and MJ):

hierarki (86%) (hierarchy)
hierarkisk (75%) (hierarchical)
harakiri (71%) (hara-kiri)
hierarkisere (67%) (sort hierarchically)
hik (67%) (hiccup)
hårlak (67%) (hair spray)
hiking (67%) (hiking)
hijacking (67%) (hijacking)
hak (67%) (notch, dent)
hagiografi (63%) (hagiography)

These options are in line with De Schryver et al. (2006) who, as mentioned above, suggest the use of various software modules to help the user and thus increase the hit rate.

De Schryver et al. (2006) mention another issue concerning lemmatisation which is specific to Swahili and other inflecting languages: the question of whether to include only word stems as is the common practice in Western lexicography, or whether to include also full word forms on the basis of a corpus. In this connection, Bergenholtz and Johnsen (2005) are criticised for being naïve in conducting their log file analysis on the basis of lemma strings only and not taking this issue into account. It should be noted, however, that the *Danish Internet Dictionary* also allows searches for morphemes, inflected forms of a lemma and random parts of words, and the various search options occur from the log files. Seeing that most users search for 'the lemma is' rather than 'the lemma begins with', 'the lemma ends with' or 'the lemma contains' (Bergenholtz and Johnsen 2005), it seems that this strategy is feasible for an agglutinative language like Danish, where the rules of word formation and grammar differ, for example, from Swahili. The lemma selection process and, consequently, the resulting log files, will necessarily to a certain extent be language-dependent as different languages require different considerations as far as dictionary compilation is concerned.

The focus of Bergenholtz and Johnsen (2005) was on improving the lemma selection on the basis of log file analysis, but other possibilities are also taken into account. It is suggested that improved search options, e.g. the possibility of searching directly in every field of the dictionary rather than just the lemma field, would result in far more detailed log files, but that the use of such detailed log files has yet to be described. This article aims to elaborate on the issue and present suggestions concerning the practical use and lexicographic relevance of such log files.

As regards the compilation of detailed log files, advanced search options have been implemented in a Danish online dictionary of music terms, the *Danish Music Dictionary*. In this dictionary, a search for a specific term will be conducted not only in the lemma field, but also in for example the definition field. The *Danish Music Dictionary* has been available on the Internet since August 2006 and was therefore not yet accessible at the time when Bergenholtz and Johnsen (2005) was written. The addition of these enhanced search options may be regarded as a further development of the log file theory. It serves a dual purpose — on the one hand, it increases the usability of the dictionary for its users, and on the other hand, it is part of a strategy for compiling more detailed, and thus more useful, log files.

The issue of MWUs was briefly touched on above, and the question of whether or not it should be possible to search for MWUs depends to some extent on the dictionary in question. It is less relevant to include MWUs in a dictionary like the *Danish Internet Dictionary* where it would be more useful to add advanced search options to enable users to search, for example, in collocations and examples, but it is highly relevant in a recently released online dictionary of Danish idiomatic expressions, the *Danish Phraseological Dictionary*. As idiomatic expressions inherently consist of more than one word, the dictionary compilers are testing a completely new approach to search options in this dictionary. The log files for the *Danish Phraseological Dictionary* will be different from lemma-oriented log files, and the initial results and implications of the strategy are described in sections 5 and 6 of this article.

3. Comparative Surveys and the Status of the Lemma

Log files are used ever more widely in the field of lexicography, not only in Denmark and South Africa, but also internationally. In Johnsen (2005), a survey of five Internet dictionaries representing different languages is carried out with the aim of establishing two facts: (1) the extent to which Internet dictionaries are used and whether the use of Internet dictionaries is increasing, and (2) whether the log files of these dictionaries show similarities in terms of user behaviour. The dictionaries in question are the above-mentioned *Danish Internet Dictionary*, *Eurodicautom* (a polylingual database compiled by translators in the European Union), *Wortschatz Deutsch* (a German and German–English dic-

tionary), *Cambridge Dictionaries Online* (various monolingual English dictionaries) and *Bokmålsordboka* (a monolingual Norwegian dictionary).

The bare figures show that all five dictionaries are widely used with an average of 6 000–6 500 daily searches in the *Danish Internet Dictionary*, 162 074 in *Eurodicautom*, 329 657 in *Wortschatz Deutsch*, 205 480 in *Cambridge Dictionaries Online* and 7 932 in *Bokmålsordboka*. The survey also reveals that the number of queries have increased over time, particularly in the first years of a given dictionary's life. In order to evaluate the user behaviour, the top 20 queries of the dictionaries as established by the log files were compared (with the exception of *Eurodicautom* as no data were available). The top 20 lists are included as Appendix 1 with English translations by HB and MJ. Interestingly, the log files for the Danish and the Norwegian dictionaries show significant similarities as both mainly contain ordinary everyday words from practically all word classes. This is, however, not the case for the English and the German dictionaries. The log files from *Wortschatz Deutsch* show that seven of the top 20 words are the names of actual persons, and a number of words related to lexicography and linguistics also appear in the list. With the exception of one adjective, the top 20 for *Cambridge Dictionaries Online* contains only nouns and verbs. The log files from *Wortschatz Deutsch* thus differ most significantly from the log files of the other three dictionaries. This may be due to the fact that users with different needs search for different items, and differences in log files from various dictionaries may thus be linked with the intended function(s) of a given dictionary. It might be imagined that the log files of a dictionary compiled for translation purposes differ from those of a dictionary intended for text production or reception. As a clear function is only stated for the *Danish Internet Dictionary*, i.e. text production, it is difficult to say whether this has any effect on the differences and similarities in the log files described above. It is, however, relevant for lexicographers to pay attention to whether dictionary users search for words that should be included in the dictionary in accordance with its function(s) or whether they might have misconceived the purpose. If the log files for the *Danish Internet Dictionary* contained as many queries for the names of actual persons as those for *Wortschatz Deutsch*, it would clearly indicate that the users in question do not understand the purpose of the dictionary. In other cases, it may be that the lexicographer has misjudged the users' needs or has been unable to communicate the dictionary's purpose clearly.

Comparative surveys and log file analyses as those described above may be interesting, and they do provide lexicographers with new knowledge, but they still do not reveal the answers to the real questions: Who are the users, why did they use the dictionary, and did they find what they were looking for? In a dictionary allowing only searches for lemmata, log file analysis has certain limitations; cf. section 1 of this article. Restricting searches to the lemma field poses two problems: what is a lemma, and what did the users do with the information they found? In order to achieve the full benefits of log file analysis, changes in the basic conception of Internet dictionaries are thus required. For

one thing, it is necessary to implement better, more advanced search options, but, more importantly, it is crucial to utilise the full potential of the electronic medium and move away from the traditional, lemma-oriented way of thinking. A structured, lemma-oriented approach is undoubtedly necessary in printed dictionaries, but not so in electronic dictionaries. Ideally, each building block of the dictionary should be entered into the database only once, and different users may choose to view the information in different ways, thus eliminating the need for a macrostructure (Bergenholtz 2005). This raises the question, then, of whether the lemma itself has also been made superfluous as every item in a dictionary may move into lemma position. In the *Danish Phraseological Dictionary* mentioned above, whatever the user chooses to enter into the search field is the lemma. The user may for instance search for the expression *stå med aben* (equivalent to the English expression: *be left holding the baby*). The dictionary then produces a list of expressions containing or relating to this idiom as it searches the entire database rather than just fields designated 'lemma'. The user may also choose to search for the exact word or phrase entered in the search field, entries containing the word(s) entered here, entries beginning with the word(s) or entries ending with the word(s). Furthermore, the user has three options regarding the purpose of the dictionary use: *hjælp til at forstå en tekst* (help in connection with understanding a text), *hjælp til at skrive en tekst* (help in connection with writing a text) and *hvis du vil vide mere* (further information). In other words, it is possible to specify or narrow down a search to exactly find the information required.

4. User Surveys

In this article, it was previously mentioned that log files may be used to collect data leading to the improvement of existing online dictionaries. This is not the only possible use of log file analyses. If one wishes to achieve new and general knowledge, it is advisable first of all to consider any kind of dictionary use, including the use of printed dictionaries. In other words, we suggest a holistic approach to user surveys according to which the purpose of conducting a user survey may be one or more of the following:

- (1) In metalexigraphic research: To present new, empirically substantiated results showing the dictionary user's need for, use of and benefits from one dictionary or a number of dictionaries.
- (2) In metalexigraphic research or a commercial survey: To test the quality of one dictionary or a number of dictionaries, possibly as part of a dictionary review or as a survey paid for by a publishing firm for the purpose of advertising and/or improving the dictionaries.
- (3) In metalexigraphic research: To collect, analyse and use data from user surveys in order to suggest concepts for new and better dictionaries.

Purposes (1) and (2) may be referred to as contemplative surveys (according to Tarp 2002). The starting point is existing dictionaries whose access methods, contents and structure are evaluated, and improvements may be suggested on the basis of the user survey. Almost all existing contributions on dictionary use of printed as well as electronic dictionaries, including log file analyses, fall under this type of survey. Purpose (3) may be referred to as a transformative survey as the purpose is not so much to repair existing dictionaries, but rather, on the basis of the analyses, to obtain arguments for future, possibly entirely new, dictionary concepts, which may not have any resemblance to the dictionaries analysed. This type of survey has not yet been carried out on a large scale. Laufer and Levitzky-Aviad (2006) have taken the first step, and this article aims to further develop the theory.

Giving priority to purpose (3) does not imply that purposes (1) and (2) should be neglected. Specifically, it is not true that user surveys conducted on the basis of printed dictionaries reveal nothing at all about dictionary use. Admittedly, compared to log file analyses, they involve only very few subjects who have carried out only a very limited number of look-ups, but unlike most existing contributions on log file analyses, they take the dictionary's various information types into account and do not focus almost solely on the lemma. In the early 1980s, the dictionary user was referred to as 'the known unknown' in several papers (Schaefer 1981: 62). This was undoubtedly the case at that time, but the question is whether the situation has changed fundamentally. In the 1980s, a large number of retrospective and introspective user surveys were carried out along the lines of Béjoint (1981) or Benbow et al. (1990); cf. the outlines in Ripfel and Wiegand (1988) and Nesi (2002: 277): *Which types of information do you need when you use a dictionary?* or *Which types of information do you consider to be important in a dictionary?* The surveys usually involved relatively few subjects, most often students, as in Benbow et al. (1990: 199):

	Daily	Weekly	Monthly	Never
Headword (e.g. for spelling)	23%	33%	25%	19%
Pronunciation	4%	16%	38%	43%
Phrases and idioms	13%	24%	37%	26%
Senses (definitions)	22%	38%	28%	12%
Illustrative quotations	12%	19%	39%	30%

Slightly, but only slightly, more realistic is the use of dictionary protocols where the user is asked to use specific dictionaries in connection with text production (Wiegand 1985), reception (Nesi 2002) or translation (Nielsen 1994: 20-32). These surveys involve very few users, and the results are thus heavily influenced by the users' individual actions. Furthermore, it is unrealistic to include bilingual dictionaries in surveys involving foreign students (as in Wiegand 1985).

Purpose (2) comprises surveys involving printed dictionaries, e.g. Benbow et al. (1990), as well as surveys involving Internet dictionaries, e.g. Ling et al. (2002), Bergenholtz and Johnsen (2005) and De Schryver et al. (2006). Surveys involving printed dictionaries have the same advantage (taking into account the various dictionary functions) and disadvantages (too few subjects, too few look-ups and completely unrealistic retrospective and introspective questions and answers) as was the case under purpose (1). Even with a relatively low number of look-ups, as in De Schryver and Joffe (2004), log files contain a much larger number of look-ups than practically all currently available surveys involving printed dictionaries. Log files of this type do not reveal any certain facts about the words, word parts or sentence combinations searched for, probably because the user often did not search for the word itself, but rather for information about the word, e.g. grammar, collocations, word formation etc., depending on the user's information needs. This may also explain the large differences between the most frequent queries in dictionaries of different languages; cf. the results in the Appendix where almost no similarities exist between the frequent queries in Danish, German, Norwegian and English, as certain words form part of numerous idiomatic expressions or involve complex grammar in some languages, but not in others. However, we have yet to find a reasonable explanation for the fact that only about half of the 128 000 entries in the *Danish Internet Dictionary* have been looked up after more than 8 million queries (Bergenholtz and Johnsen 2005). These data are entirely different from the data from three other online dictionaries of technical terms. In the *Danish Music Dictionary*, the *Danish-English Accounting Dictionary* and the *English-Danish Accounting Dictionary* (with approximately 4 000, 5 000 and 6 000 entries, respectively), almost all entries have been looked up after a much lower number of queries. This does not mean that existing log file analyses are of no value — they may be used to reveal lemma lacunae or improve the search options.

Purpose (2) also comprises more sophisticated log file analyses such as those presented by Ling et al. (2001 and 2002). These log files contain information not only about searches for lemmata and parts thereof, but also information based on search strings, e.g. full-text searches and searches using synonyms and conceptually associated words:

A query pattern represents a set of user queries with the same or similar intention, and thus is associated with an article as the answer. It is difficult to capture the "similar intention" in natural language, and therefore, we use both syntactic constraint (such as stemming of keywords [...] as well as semantic constraints (such as generalized concepts and synonyms [...]) in our definition of queries with "similar intentions". (Ling et al. 2002: 1103)

The purpose of Ling et al. (2002) is to optimise the user's search options in the *Microsoft Encarta Encyclopaedia* by establishing patterns of associative search words for *Encarta's* 42 000 entries. Furthermore, the aim is to find particularly popular themes in order to add more entries related to these themes. These log

file analyses cannot rightly be characterised as a true description of the dictionary users' actions. The collected data are so general that they may at best be used to improve the search modalities, but not the lexicographic content. This is for example the case when, in their analysis of 4.8 million queries, Ling et al. (2002: 1103) conclude that 52.5% of the queries are carried out using one search word, 32.5% using two search words, 10% using three search words and 5% using four search words or more, which, in our opinion, is of no direct relevance to the lexicographic concept. Moreover, when analysing a much smaller number of queries (271 803), Ling et al. (2001) reach the same conclusion, i.e. that exactly 52% of the queries are carried out using one search word. These analyses show but one fact: that an online dictionary allowing the users to search by means of one search word only does not correspond to the users' habits or needs.

An interesting example of purpose (2) is Laufer and Hill (2000), who allow the user to compose an individual search profile:

If, for example, the learner is interested in a quick L2–L1-translation, the option should be available. If, on the other hand, s/he is interested in examples of usage, or in grammatical information, or in a definition, each type of information should be accessible via another lookup option. Log files would record which of these options were selected for which words. (Laufer and Hill 2000: 59)

This suggestion has possibilities, but its consequences have not been given full consideration. If a user wants to understand a word or a collocation in a text, the user is only 'interested' in the equivalent and nothing else. Information on grammar or collocations, on the other hand, will help the user solve a problem in connection with L2 text production. Finally, the user may be 'interested' in all types of information if he/she wants to learn as much as possible in a cognitive process independent of text-related communicative problems. This test, as well as Laufer and Levitzky (2006), involves too few subjects to provide generally reliable indications of dictionary use. 75 students participated in a constructed reception test, not a test in a real life setting. In each case, the students were asked to select a search profile (Laufer and Levitzky-Aviad 2006: 150f) and given eight possible combinations to choose from. For text production in a foreign language, the students had the following preferences (listed according to the overall selection of all the students):

Translation + Definitions + Examples
English translations only
Translation + Definitions
Translation + Examples
Translation + Definitions + Examples + Thesaurus

The last three combinations were used only to a very limited extent:

Translation + Examples + Thesaurus
 Translation + Examples + Thesaurus
 Translation + Definitions + Thesaurus

We believe that these log files provide an unreliable and uncertain picture, as only the very experienced user (experienced in the use of the dictionary in question) will be able to choose among the many combinations correctly and in accordance with the required function(s). As we see it, a controlled process for inexperienced dictionary users and extensive freedom of choice for experienced users would enable lexicographers to document dictionary use in a way that would really be useful to transformative lexicography.

5. Suggested Structure of Log Files for Internet Dictionaries

If log files really are to reflect how and for which purpose dictionaries are used, a function-oriented set of search and link options is required. Surely, this is not to say that all users should be presented with as many lexicographic data as possible, or even with more data than they desire:

I will be shamelessly selfish and ask for the impossible. I will advocate for a dictionary that will always adapt to my needs and always be ready to provide me with exactly the answer that I need and will also agree with. I also expect the dictionary to be able to give me satisfactory answers to those questions that I forget to ask. (Varantola 2002: 31)

On the contrary, all present experience with searches on the Internet suggests that the main problem usually involves receiving too many hits or texts that are too long. Users of Internet dictionaries may also be inundated by too much information. Another term used in connection with dictionary searches is 'over generalization' (Ling et al. 2002: 1105). In the mind of the user, the perfect dictionary provides just the amount of information required to fulfil the user's need. If a user for example needs assistance in translating a collocation, this collocation will be the search string entered by the user, and the translation will be exactly the short and exhaustive answer the user needs. Translated into the terminology used in connection with printed dictionaries, the user's search string is for instance the lemma **break a gene**, and this lemma and its corresponding equivalent constitute the dictionary entry, i.e. in an English–Danish dictionary **break a gene** *skære et gen over*. In the ideal situation, a one-to-one relation occurs, but most often a one-to-many relation exists between search strings and the entries shown. This type of dictionary does not contain a fixed number of entries, but rather all potential entries permitted by the dictionary database on the basis of the search criteria. Few, if any, Internet dictionaries function like this yet (but they will become reality within the next two or three years), and the user is merely offered ready-made entries like those found in printed dictionaries.

It is not, however, futuristic speculation to say that the current dictionary typologies of monolingual, bilingual and polylingual dictionaries or language dictionaries and subject dictionaries are relevant classifications when it comes to designing lexicographically relevant log files. In fact, the various user surveys conducted so far, and also the majority of log file analyses, are based on the assumption that a dictionary is a dictionary without any kind of distinction between dictionary types. They have, to a certain degree, distinguished between monolingual and bilingual dictionaries, but this is surely not the only suitable distinction to be made, as users very much need a combination of dictionary types taking into account the various needs of the users; cf. Mugdan (1992) and Laufer and Levitzky-Aviad (2006). This line of thought is not a new one:

More original typologies are undoubtedly imaginable, for instance one that would be based on the functions of dictionaries and/or on the different types of organization of addresses (that is, types of organization of access to information). (Hausmann et al. 1989: xix)

The idea has been transformed into suggestions for a concrete dictionary typology in Bergenholtz and Kaufmann (1997: 100), who distinguish between reference books for needs relating to text-dependent and text-independent problems. Today, the use of the terms 'dictionaries for communicative functions' and 'dictionaries for cognitive functions' are preferred to refer to dictionaries designed to assist the user in resolving issues relating to text reception, text production and translation on the one hand and dictionaries designed to assist the user in resolving issues relating to the acquisition of knowledge on the other. In the first scenario, the user has a problem related to a text that he/she does not understand, has difficulties formulating or has difficulties translating. In the second scenario, the user has a general need for new knowledge, be it specific knowledge about a given word or expression or general knowledge about a particular subject. With this theory as a starting point, lexicographers may begin to comprehend how a dictionary is really used. Existing contributions on the subject, such as De Schryver et al. (2006), may present many figures, but their interpretational value to lexicography is limited.

These considerations lead to the following suggestions for the collection of log file data¹:

- (1) the user's IP address,
- (2) the user's browser type,
- (3) the user situation,
- (4) which fields the user searches in if he/she chooses to define his/her needs according to the user situation as such,
- (5) the user's search string,
- (6) the search criteria,
- (7) the date or time of the search,
- (8) the result of the search, including all the entries that were found,

- (9) which links the user clicks on in the entry found (entry being what is shown on the screen),
- (10) which fields the user copies from, and
- (11) which outside matter the user consults.

Not all data are of lexicographic relevance. The activities of search robots, for instance, are not relevant, yet they account for 30% of all searches in the *Danish Music Dictionary* over the first five months. They have IP addresses starting with 65 or 66 and should be filtered out. Log file item (10) cannot be logged in dictionaries using HTML, which is a so-called stateless language and therefore only allows for the logging of input commands initiated by the user. The most important items among the log file data listed above are (3)-(6) and (9)-(11). Item (5) is the search method used by the computer (the search engine). It is not just a question of defining the search string as 'is', 'contains', 'begins with' or 'ends with', but also of the algorithms used in the programming. Of particular interest is log file item (4), which will show the fields the user believes to be necessary in the basic user situation. We believe that the user will most often choose the field containing the lexicographic definition for reception problems, and in the case of text production problems, we assume that the user will also choose fields containing synonyms and antonyms, grammatical information, collocations and examples. In a translation situation, the user's choice in the case of L1 problems will most likely be the same as for reception problems, and in the case of L2 problems, it will probably be the same as for text production problems with the addition of the equivalent field. Finally, users who are on a quest for knowledge in the widest sense of the word are likely to require all information, including information on etymology and Internet references. Log file item (11) is also interesting. A recurring argument in dictionary related discussions is that a dictionary's outside matter, particularly its user instructions, are hardly ever consulted. This argument is put forward despite the fact that the only major survey known to us (Wolf 1994) shows that over 50% of all users have read the user instructions within the first two years of purchasing a dictionary. Our first log file data prove that Wolf's results concerning the use of outside matter in printed dictionaries are transferable to the use of outside matter in online dictionaries. The user clicks on the outside matter in more than 5% of all searches. Since every user on average conducts two searches a day, it means that 10% of the users consult the outside matter. The following figures (indicating the number of look-ups) are log file data from the *Danish Phraseological Dictionary* from the period 13 March to 26 March 2007:

<i>om ordbogen</i> (a kind of preface)	73
<i>søgetips</i> (short user instructions)	51
<i>brugervejledning</i> (more detailed user instructions)	46
<i>litteratur</i> (relevant literature)	31
<i>kontakt</i> (information on how to contact the editors)	28
<i>copyright</i> (information on dictionary rights)	22

These figures are to be viewed in relation to the number of dictionary searches conducted in the same period of time (13 March to 26 March 2007):

dictionary searches	4 010	94.11%
outside matter	251	5.89%
total	4 261	100.00%

The user's basic need for help, i.e. log file item (3), is also important. We suggest that this part be made mandatory, which means that the user will not be able to continue the search unless he/she makes a choice. Alternatively, instead of forcing the user to actively choose, one of the functions may be used as the default option in case the user does not make a choice. This would, however, produce figures that are not entirely reliable as the default button may lead to a biased result, seeing that many users might not take the trouble to make an active choice. Nonetheless, we have opted for a default button for the time being in our initial experiments with log files of this type for the purpose of user friendliness.

6. First Experiments with Function-oriented Log Files

We have used two dictionaries to conduct our initial experiments with the logging of dictionary functions. It seems that using a default option is of no major consequence. In the case of the *Danish Music Dictionary*, the default option was initially *help in connection with understanding a text*, but this was later changed to *further information*. Yet, the differences in the log file data before and after this change are relatively small:

Searches from 9 November 2006 to 4 December 2006:

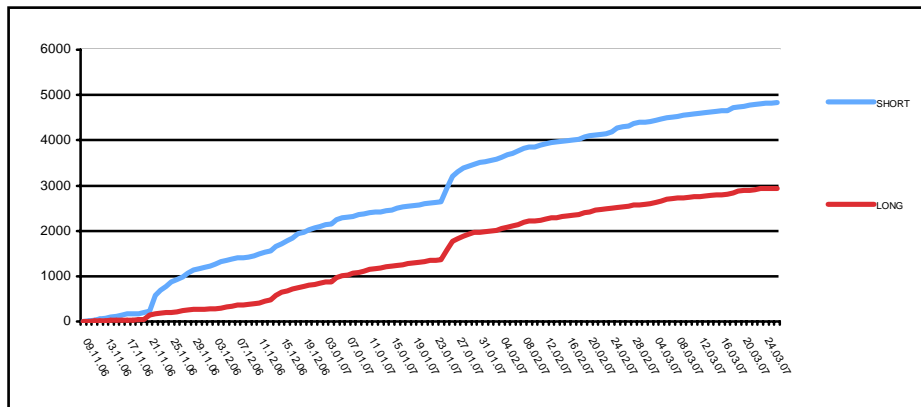
understanding a text	1 979	54.38%
further information	1 660	45.62%
total	3 639	100.00%

Searches from 5 December 2006 to 26 March 2007:

understanding a text	8 733	53.77%
further information	7 507	46.23%
total	16 240	100.00%

In other words, these figures indicate that the use of a default option does not measurably influence the result. However, it is interesting to observe that experienced users apparently to an increasing degree just 'settle for' the information they need, i.e. assistance in relation to text reception problems. Furthermore, the curves illustrating the two search options move ever further apart over time. This leads to the conclusion that an experienced user requiring help in a reception situation has learned that he/she will get a clear answer to

his/her question by actively choosing the reception button rather than making use of the default option:



The *Danish Music Dictionary* is, as can be deduced from the above description of its log files, a polyfunctional dictionary with two functions:

- (1) help in solving reception problems, which is indicated by the button *help in connection with understanding a text*, and
- (2) help in the acquisition or expansion of knowledge, which is indicated by the button *further information*.

When using the first button, the user is presented with a short lexicographic definition, whereas clicking the other button provides the user with a detailed definition including illustrations, note examples and links to a short, integrated presentation of music theory. It was our assumption that most users would choose the detailed version. As can be seen from the log file data for both default options and a total of 19 881 searches, this is not the case, neither when the short definition, nor when the long definition for the acquisition of knowledge is used as default option:

understanding a text	10 714	53.89%
further information	9 167	46.11%
total	19 881	100.00%

Further statistics of the users who move from the reception entry to the knowledge acquisition entry, i.e. from the short entry to the long entry, or vice versa by clicking the internal link may be deduced from the figures. The remaining searches would reveal in how many instances the user's need for information was fulfilled by the first search. Such statistics are not available for the experimental dictionaries that we have compiled so far. Still, the log files reveal a crucial point. The initial log file results alone disprove the theories of Wierzbicka (1985) with her proposals for excessively long lexicographic definitions. Such

detailed definitions are required when cognitive functions are involved whereas short definitions are more appropriate in relation to communicative functions.

Similar results appear from the log files of the *Danish Phraseological Dictionary* where a log file system has been in operation since the end of February 2007. This dictionary has three functions, two communicative functions and one cognitive function, which are indicated by the following buttons: (1) help in connection with understanding a text, (2) help in connection with writing a text, and (3) further information, i.e. knowledge acquisition. The first log files for this dictionary show that a large majority of its users only want information on (1) the meaning of a phrase, whereas (2) help in connection with text production, which includes further grammatical information, synonyms, collocations and examples, and (3) as much information about an idiom or a proverb as possible, which, in addition to the information provided under (2), includes etymology and Internet references to relevant contributions on phraseology, are less popular:

understanding a text	9 483	58.01%
writing a text	4 320	26.43%
further information	2 543	15.56%
total	16 346	100.00%

As was the case with the *Danish Music Dictionary*, we expected that a clear majority of the users would choose option (3), but we have had to adjust our expectations. The default option in this dictionary is help in connection with reception, and the figures should therefore be interpreted with a certain amount of caution. Further log file data will enable us to determine how many users move from the default search option to one of the other two options, but the differences between the three functions are considerable with more than half of the searches being conducted for reception purposes (which may indicate that the default option influences the result to some degree). Yet, it is interesting that more users choose option (2) than option (3), which leads us to conclude that more is not always better in the mind of the dictionary user. Removing the default option completely and thus forcing the users to make an active choice in order to conduct a search in the dictionary would therefore provide a more realistic picture.

When conducting these searches, the user cannot choose which types of information he/she wants to see. We believe that such a controlled search process is the most advantageous option for users who (just) want an answer to a specific question. For experienced users or users who are adept at experimenting and wish to do so, the possibility of choosing one of the three buttons and defining which fields to be shown should be implemented. There is no way of knowing for sure how the above figures would change if this possibility is realised. Our guess is that less than 10% of the users would bother. Intelligent

log files of this type have yet to be analysed on a large scale. It is our belief that analyses of this kind would bring about new and constructive metalexigraphic knowledge, particularly if used in conjunction with analyses of the large number of e-mails received by compilers of Internet dictionaries providing an e-mail address for user inquiries.

7. Moral

User surveys will never provide a 'true' picture without any limitations and conditions. Compared to the present guessing competitions and to lemma-oriented log file analyses, analyses of function-oriented log files are much closer to the truth, especially if those searches where the user has moved from one function to another are filtered out. *Theory without empirical data is empty*. This maxim is applicable to many, if not most, surveys concerning printed dictionaries in which the empirical basis is, at best, rather weak. *Empirical data without theory are empty*. This maxim may, in part, be applied to existing log file analyses concerning the use of Internet dictionaries. Many figures are presented, but their lexicographic relevance is limited. The aim of this article is to contribute to such a theory as it is our conviction that function oriented log file analyses are, for the time being, the best and most reliable way to look over the dictionary user's shoulder.

Endnote

- † Some of these suggestions were put forward by Richard Almind, Aarhus School of Business, Aarhus, Denmark. We would like to thank him for several useful and constructive discussions.

References

- Béjoint, Henri.** 1981. The Foreign Student's Use of Monolingual English Dictionaries: A Study of Language Needs and Reference Skills. *Applied Linguistics* 2: 207-222.
- Benbow, Timothy, Peter Carrington, Gayle Johannesen, Frank Tompa and Edmund Weiner.** 1990. Report on the *NEW Oxford English Dictionary* User Research. *International Journal of Lexicography* 3(3): 155-203.
- Bergenholtz, Henning.** 2005. Den usynlige elektroniske productions- og korrekturordbog. *Lexico-Nordica* 12: 19-38.
- Bergenholtz, Henning and Mia Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes, Journal of Linguistics* 34: 117-141.
- Bergenholtz, Henning and Uwe Kaufmann.** 1997. Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field. *Hermes* 18: 91-125.
- Bokmålsordboka* = Wangenstein, Boye (Ed.). *Bokmålsordboka — definisjons- og rettskrivningsordbok*. <http://www.dokpro.uio.no/ordboksoek.html>.

Cambridge Dictionaries Online. <http://dictionary.cambridge.org>.

Danish–English Accounting Dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz in cooperation with Mads Melgaard, Trine Middelboe, Brit Sørensen, Mia Johnsen, Rie Bobjerg Nielsen, Jóna Ellendersen, Amalie Kofoed Stender and Vibeke Vrang. 2004–2007. *Den Engelsk–Danske Regnskabsordbog/English–Danish Dictionary of Accounting*. Database and design: Richard Almind. Implementation and encoding of web pages: Caspar Thomsen. <http://www.regnskabsordbogen.dk/regn/dkgb/dkgbregn.aspx>.

Danish Internet Dictionary = Henning Bergenholtz and Vibeke Vrang in cooperation with Lena Lund, Helle Grønberg, Maria Bruun Jensen, Signe Rixen Larsen, Rikke Refslund, Mia Johnsen, Katja Å. Laursen, Sophie Leegaard and Maj H. Bukhave. 2002–2007. *Den Danske Netordbog*. Database and design: Richard Almind. <http://www.ordbogen.com/ordbog/ddno/>.

Danish Music Dictionary = Inger Bergenholtz in cooperation with Richard Almind and Henning Bergenholtz. 2006. *Musikordbogen*. <http://www.musikordbogen.dk>.

Danish Phraseological Dictionary = Henning Bergenholtz and Vibeke Vrang. 2007. *Ordbogen over Faste Vendinger*. Database and design: Richard Almind. <http://www.idiomordbogen.dk>.

De Schryver, Gilles-Maurice and David Joffe. 2004. On How Electronic Dictionaries are Really Used. Williams, G. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 187–196. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

De Schryver, Gilles-Maurice, David Joffe, Pitta Joffe and Sarah Hillewaert. 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67–83.

English–Danish Accounting Dictionary = Sandro Nielsen, Lise Mourier and Henning Bergenholtz in cooperation with Mia Johnsen, Rie Bobjerg Nielsen, Jóna Ellendersen, Amalie Kofoed Stender and Vibeke Vrang. 2006–2007. *Den Engelsk–Danske Regnskabsordbog/English–Danish Dictionary of Accounting*. Database and design: Richard Almind. Implementation and encoding of web pages: Caspar Thomsen. <http://www.regnskabsordbogen.dk/iasgbdk>.

Eurodicautom. <http://europa.eu.int/eurodicautom/Controller>.

Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta. 1989. Preface. Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta (Eds.) 1989. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*. First Volume: xvi–xxiv. Berlin/New York: De Gruyter.

Johnsen, Mia Steen. 2005. Logfiler som leksikografisk analyseinstrument og hjælpeværktøj. https://merkur2.asb.dk/F/C5NCNGIFXTRQS62JNBMTJJ8X22522SV62KHICN6YVYJQQXBEGM-00166?func=service&doc_library=HBA01&doc_number=000139835&line_number=0001&service_type=MEDIA.

Laufer, Batia and Monica Hill. 2000. What Lexical Information Do L2 Learners Select in a Call Dictionary and How Does It Affect Word Retention? *Language Learning & Technology* 3(2): 58–76.

Laufer, Batia and Tamar Levitzky-Aviad. 2006. Examining the Effectiveness of 'Bilingual Dictionary Plus' a Dictionary for Production in a Foreign Language. *International Journal of Lexicography* 19(2): 135–155.

Ling, Charles X., Jianfeng Gao, Huajie Zhang, Weining Qian and Hongjiang Zhang. 2001. Min-

- ing Generalized Query Patterns from Web logs. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34) Volume 5*. Washington, DC: IEEE Computer Society Press 5020. <http://research.microsoft.com/~jfgao/paper/hicss01.pdf> (February 2007).
- Ling, Charles X., Jianfeng Gao, Huajie Zhang, Weining Qian and Hongjiang Zhang.** 2002. Improving Encarta Search Engine Performance by Mining User Logs. *International Journal of Pattern Recognition and Artificial Intelligence* 16(8): 1101-1116.
- Mugdan, Joachim.** 1992. On the Typology of Bilingual Dictionaries. Hyldgaard-Jensen, Karl and Arne Zettersten (Eds.). 1992. *Symposium on Lexicography V. Proceedings of the Fifth International Symposium on Lexicography May 3–5, 1990, at the University of Copenhagen*: 17-24. Tübingen: Niemeyer.
- Nesi, Hilary.** 2002. A Study of Dictionary Use by International Students at a British University. *International Journal of Lexicography* 15(4): 277-305.
- Nielsen, Sandro.** 1994. *The Bilingual LSP Dictionary. Principles and Practice for Legal Language*. Tübingen: Narr.
- Ripfel, Martha and Herbert Ernst Wiegand.** 1988. Wörterbuchbenutzungsforschung. Ein kritischer Bericht. Wiegand, Herbert Ernst (Ed.). 1988. *Studien zur neuhochdeutschen Lexikographie VI, 2. Teilband*. Germanistische Linguistik 87-90: 491-520. Hildesheim: Olms Verlag.
- Schaeder, Burkhard.** 1981. *Lexikographie als Praxis und Theorie*. Tübingen: Niemeyer.
- Tarp, Sven.** 2002. Translation Dictionaries and Bilingual Dictionaires — Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Varantola, Krista.** 2002. Use and Usability of Dictionaries: Common Sense and Context Sensibility. Corréad, M.-H. (Ed.). *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*: 30-44. Göteborg: EURALEX.
- Wiegand, Herbert Ernst.** 1985. Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzer einsprachiger Wörterbücher. Bergenholtz, Henning and Joachim Mugdan (Eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.–30.6.1984*: 20-98. Tübingen: Niemeyer.
- Wierzbicka, Anna.** 1985. *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.
- Wolf, Birgit.** 1994. Wörterbuch und Benutzer — Versuch einer empirischen Untersuchung. ed. by Brauße, Ursula and Dieter Viehweger (Eds.). *Lexikonomie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis*: 295-389. Tübingen: Niemeyer.
- Wortschatz Deutsch* = Quasthoff, Uwe, Stefan Bordag, Christian Biemann and Matthias Richter in cooperation with Christian Wolff, Fabian Schmidt, Karsten Böhm, Martin Läuter, Timo Böhme, Jens Drawehn, Frank Fischer, Sandra Liebold, Katja Mannekens and Martin Quedsted. 1998–2005. *Wortschatz Lexicon*. <http://wortschatz.uni-leipzig.de>.

Appendix: Top 20 queries in four different Internet dictionaries

The Danish Internet Dictionary (2003–2004)

- 1 gå (walk)
- 2 a (a, each)
- 3 hest (horse)
- 4 for (to, for)
- 5 ad (by, along)
- 6 arbejde (work)
- 7 kompetence (competence)
- 8 hus (house)
- 9 finke (finch)
- 10 tage (take)
- 11 bil (car)
- 12 empati (empathy)
- 13 tid (time)
- 14 kognitiv (cognitive)
- 15 indenfor (in, inside)
- 16 i (in)
- 17 vand (water)
- 18 godt (good, well)
- 19 linie (line)
- 20 hund (dog)

Bokmålsordboka (2000–2005)

- 1 feil (mistake, error)
- 2 dessverre (unfortunately)
- 3 verken (neither)
- 4 fitte (pussy)
- 5 hverken (neither)
- 6 interessant (interesting)
- 7 desverre (unfortunately)
- 8 internett (Internet)
- 9 engelsk (English)
- 10 interesse (interest)
- 11 kognitiv (cognitive)
- 12 hei (hi)
- 13 interessert (interested)
- 14 pragmatisk (pragmatic)
- 15 empati (empathy)
- 16 patetisk (pathetic)
- 17 tunnel (tunnel)
- 17 enda (even)
- 18 ordbok (dictionary)
- 20 nysgjerrig (curious)

Wortschatz Deutsch (February 2005)

- 1 synonyme (synonyms)
- 2 lexikon (lexicon)
- 3 wortschatz (vocabulary)
- 4 deutsch (German)
- 5 wortschatz deutsch (the name of the dictionary)
- 6 serena williams
- 7 michael jackson
- 8 britney spears
- 9 deutsches wörterbuch (German dictionary)
- 10 synonymwörterbuch (thesaurus)
- 11 lance armstrong
- 12 nicole kidman
- 13 wortschatz leipzig (vocabulary leipzig)
- 14 george bush
- 15 thesaurus
- 16 fremdwörter (loanwords)
- 17 wortschatz uni leipzig (vocabulary university leipzig)
- 18 dudn (the name of a German reference work)
- 19 saddam hussein
- 20 deutsch wörterbuch (German dictionary)

Cambridge Dictionaries Online (2004)

- 1 advice
- 2 liaison
- 3 effect
- 4 regard
- 5 comply
- 6 appreciate
- 7 commit
- 8 assess
- 9 endeavour
- 10 acquire
- 11 paradigm
- 12 information
- 13 analyse
- 14 intend
- 15 affect
- 16 provide
- 17 idiom
- 18 propose
- 19 emphasize
- 20 ubiquitous

Professionele vertalers as woordeboekgebruikers*

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: 'n Ondersoek na die eienskappe en werkswyse van professionele vertalers en die kenmerke van die vertaalproses toon dat professionele vertalers spesifieke leksikografiese behoeftes het wat nie volledig deur tradisionele tweetalige woordeboeke bevredig kan word nie. Die probleme of behoeftes wat by die professionele vertaling van T2 > T1-tekste kan ontstaan deur die leksikografiese data in 'n vertaalwoordeboek opgelos of bevredig kan word, word ondersoek. Die begrippe *tweetalige woordeboek*, *vertalende woordeboek* en *vertaalwoordeboek* word bespreek. Op grond van die werklike doel en funksies van 'n vertaalwoordeboek word oorwegings uitgelig wat by die produksie van 'n vertaalwoordeboek in ag geneem moet word.

Slutelwoorde: VERTAALPROSES, VERTAALVERMOË, VERTAALWOORDEBOEK, TWEETALIGHEID, TWEETALIGE LEKSIKOGRAFIE, VERTAALEKWIVALENTE, VERTALENDE WOORDEBOEK, WOORDEBOEKFUNKSIES, TEKSPRODUKSIE, TEKSRESEPSIE

Abstract: Professional translators as dictionary users. An examination of the characteristics and procedures of professional translators and the characteristics of the translation process indicate that professional translators have specific lexicographic needs that traditional bilingual dictionaries do not completely satisfy. Problems or needs that may arise during the professional translation of T2 > T1 texts and may be solved or satisfied by lexicographic data in a translation dictionary are examined. The concepts *bilingual dictionary*, *translating dictionary* and *translation dictionary* are discussed. On the basis of the real purpose and functions of a translation dictionary, considerations that are singled out need to be considered in the production of a translation dictionary.

Keywords: TRANSLATION PROCESS, TRANSLATION COMPETENCE, TRANSLATION SKILLS, TRANSLATION DICTIONARY, BILINGUAL LEXICOGRAPHY, BILINGUALISM, TRANSLATION EQUIVALENTS, LEXICOGRAPHIC FUNCTIONS, TEXT PRODUCTION, TEXT RECEPTION

1. Inleiding

Vertalers het dikwels ambivalente gevoelens oor tweetalige woordeboeke. Enersyds beskou hulle dit as 'n onontbeerlike hulpmiddel, en andersyds kry

* 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.

hulle in die praktyk soms min praktiese nut daaruit — hul tweetalige woordeboek bied óf geen hulp nie óf slegs gebrekkige leiding. Daar is selfs vertalers wat beweer dat hulle 'n tweetalige woordeboek so ver moontlik vermy (Roberts 1990: 74).

In hierdie artikel word die probleem aan die hand van vroeë soos die volgende ondersoek: Hoekom is tweetalige woordeboeke nie vir professionele vertalers 'n voldoende hulpmiddel nie? Hoe verskil 'n tweetalige woordeboek, 'n vertalende woordeboek en 'n vertaalwoordeboek? Watter vereistes stel professionele vertalers aan 'n vertaalwoordeboek? Hoe kan tweetalige woordeboeke aangepas word om vir vertalers 'n beter hulpmiddel te wees?

2. Die werklike doel van 'n woordeboek

Woordeboeke het oorspronklik in antwoord op die praktiese behoeftes van bepaalde gebruikers ontstaan. Wiegand (1988; vgl. Tarp 2000: 193, en Bergenholtz en Tarp 2003: 179) bring in hierdie verband die gedagte van die "werklike doel" van 'n leksikografiese naslaanwerk ter sprake, wanneer hy na leksikografiese naslaanwerke verwys as gebruiksprodukte wat vir 'n spesifieke doel geproduseer word. Die werklike doel van 'n leksikografiese naslaanwerk is volgens hom dat die gebruiker dit kan benut om inligting te onttrek uit die leksikografiese data wat die betrokke werk op 'n bepaalde naslaangebied verskaf.

Tarp (2000: 193) stem saam dat dit 'n akkurate beskrywing van die werklike doel van 'n woordeboek is, maar meen dié definisie is baie algemeen. Indien woordeboeke erken word as gebruiksprodukte wat saamgestel word om in die besondere behoeftes van bepaalde gebruikers te voorsien, is dit noodsaaklik om telkens die lede van 'n bepaalde gebruikersgroep se eienskappe en behoeftes te ontleed ten einde die leksikografiese funksies van die betrokke leksikografiese naslaanwerk te bepaal (Tarp 2000: 196).

Die leksikograaf se eerste taak is dus om 'n profiel van die beoogde gebruikersgroep en 'n tipologie van beoogde gebruikersituasies saam te stel en dan vas te stel waar probleme of behoeftes kan ontstaan wat deur die leksikografiese data in 'n woordeboek opgelos of bevredig kan word. Op grond hiervan kan die funksies en werklike doel van die beoogde woordeboek dan bepaal word (Bergenholtz en Tarp 2003: 173).

3. Vertalers as woordeboekgebruikers

Dis belangrik om helderheid te verkry oor wie met "vertalers" bedoel word en wat "vertaling" behels, sodat vasgestel kan word watter vertaalprobleme in die vertaalproses kan ontstaan en hoe die woordeboek moet lyk wat die vertaler kan help om sodanige probleme op te los. Die vertaling wat 'n professionele vertaler doen, verskil van ander aktiwiteite waarna in die metaleksikografie as "vertaling" verwys word. Dit verskil byvoorbeeld van die vertaling wat die leksikograaf tradisioneel doen by die saamstel van 'n tweetalige woordeboek, of

van die vertaling wat as 'n oefening in die taalklas gedoen word, of van die vertaling wat die toeris met behulp van 'n fraseboek in 'n vreemde land doen.

Vir die doeleindes van hierdie bespreking word professionele Afrikaanssprekende vertalers van nieliterêre tekste uit Engels in Afrikaans ($T2 > T1$) as die primêre gebruikersgroep geneem. Die gebruikersituasie wat ondersoek word, het betrekking op die proses wanneer 'n bronteks uit Engels ($T2$) in Afrikaans ($T1$) vertaal word.

Moontlike sekondêre gebruikers is ervare professionele vertalers wat literêre tekste uit Engels in Afrikaans ($T2 > T1$) vertaal; onervare professionele vertalers (minder as drie jaar in die praktyk) wat literêre en nieliterêre tekste uit Engels in Afrikaans ($T2 > T1$) vertaal; en semiprofessionele vertalers wat literêre of nieliterêre tekste uit Engels in Afrikaans ($T2 > T1$) vertaal.

Die vertaling van literêre tekste word as 'n potensiele gebruikersituasie bygevoeg, omdat literêre en nieliterêre vertaling nie noodwendig in die praktyk van mekaar geskei word nie. Nieliterêre tekste (veral reklametekste en joernalistieke tekste) stel dikwels aan die skeppende vermoëns van vertalers dieselfde eise as literêre tekste.

Onervare en semiprofessionele vertalers kan beskryf word as vertalers wat nog deur 'n steil leerkurwe vorder. Die vertaler staan byvoorbeeld aan die begin van sy of haar loopbaan of is 'n student in vertaling en ontvang intensiewe opleiding vir die vertaalberoep, hetsy aan 'n opleidingsinstansie of deur middel van indiensopleiding.

3.1 Die profiel van 'n vertaler

Volgens Neubert (2000: 6) vorm die volgende vyf vermoëns die parameters van vertaalvermoë of vertalingsbevoegdheid: taalvermoë, teksvermoë, vakvermoë, kultuurvermoë en oordragvermoë. Hierby kan probleemoplossingsvermoë as die sesde vermoë genoem word waarvoor die professionele vertaler beskik.

3.1.1 Taalvermoë

Die professionele vertaler het naastenby volmaakte beheer oor die grammatika en woordeskat van sowel die brontaal ($T2$) as die doeltaal ($T1$) (Neubert 2000: 7). Ofskoon dit op sigself nie waarborg dat iemand ook 'n bedrewe vertaler sal wees nie, is volkome tweetaligheid¹ 'n voorvereiste vir vertalingsbevoegdheid.

In haar ondersoek na die verband tussen tweetaligheid en vertaalvermoë konsentreer Presas (2000: 21) op twee sake: (a) watter vaardighede die tweetalige persoon in elk van sy/haar twee tale bemeester en (b) wat die kognitiewe effek van hierdie leerproses is (hoe die struktuur van die tweetalige geheue lyk en wat die twee eienskappe behels wat met tweetalige taalgebruik geassosieer word, te wete taalverwarring en kodewisseling).

Eerstens word algemeen aanvaar dat tweetalige taalgebruikers twee tale verstaan en hulle in twee tale kan uitdruk. Dit beteken resepsie én produksie is

ter sprake — by vertaling spesifiek teksresepsie en teksproduksie. Die vaardighede wat by direkte vertaling ($T2 > T1$) betrokke is, is teksresepsie in T2 en teksproduksie in T1.

Tweedens stel Presas ondersoek in na die kognitiewe meganismes wat vertaling onderlê. Beskik die tweetalige persoon oor 'n onafhanklike taalgeheue vir elk van sy/haar twee tale, of 'n interafhanklike taalgeheue? Dit gaan dus oor "the relationship between the linguistic signs of each one of the languages and the mental representations which the speaker associates with these signs" (Presas 2000: 22).

Volgens Presas toon psigolinguistiese navorsing dat die aktivering van gedagtebeelde ("mental images") en taalgeheue 'n wesenlike komponent van die resepsie en produksie van taal is. Hieruit volg dat die tipe tweetalige taalgeheue wat die vertaler openbaar, noodwendig tydens vertaling 'n bepaalde invloed op die proses van teksresepsie en teksproduksie sal hê. Presas verdeel vertalers in vier tipes na gelang van hul tweetalige taalgeheue en die resepsie-produksie-proses wat hulle op grond daarvan volg:

- Vertalers met 'n *assosiatiewe* werkswyse verbind die leksikale elemente of tekseienskappe van die een taal met leksikale elemente of tekseienskappe van die ander taal.
- Vertalers met 'n *bykomstige* of *ondergeskikte* werkswyse heg gedagte-inhoud aan net een van die twee tale.
- Vertalers met 'n *gemengde* werkswyse assosieer leksikale elemente van een taal met 'n enkele repertorium van gedagte-inhoud, waaruit hulle ook assosiasies met leksikale elemente van die ander taal haal.
- Vertalers met 'n *gekoördineerde* werkswyse assosieer die leksikale elemente van hul eerste taal met 'n eie repertorium van gedagtebeelde en assosieer die spesifieke gedagte-inhoud van dié eerste repertorium met spesifieke gedagte-inhoud van 'n tweede repertorium, wat dan weer met leksikale items van die tweede taal in verband gebring word.

Presas (2000: 25) meen dis moontlik dat enige vertaler die eienskappe van meer as een van die vier tipes vertoon, of dat vertalers in verskillende stadiums van opleiding die dominante kenmerke van 'n bepaalde tipe sal vertoon. Sy reken egter dat 'n gekoördineerde werkswyse die ideale vertalingsbevoegdheid verteenwoordig.

Taalverwarring ("interference") en kodewisseling ("code-switching") is twee eienskappe wat met tweetaligheid geassosieer word. Hoewel die professionele vertaler nie in dieselfde mate as onervare vertalers aan die tussentrede van T1 by die resepsie van T2 onderhewig is nie — dis waarskynlik wat die verskynsel van vals vriende onderlê — kom taalverwarring voor wanneer onder-na-bo-prosesserings tydens die resepsieproses nie voltooi word deur vol-

doende bo-na-onderprosessering nie (Presas 2000: 25). Tweedens oefen die T2-bronteks dikwels by die produksie van T1 'n "hipnotiese mag" oor die vertaler uit wat volgens Presas waarskynlik te wyte is aan "the prevalence of mechanisms of the associative type at the lexical level, which preclude the relevant stage of mental content analysis". 'n Moontlike oplossing vir beide tipes taalverwarring is dat vertalers die resepsie-produksie-meganismes van die gekoördineerde vertaler moet verwerf (Presas 2000: 26) — 'n proses wat klaarblyklik bevorder kan word as eentalige verklarende woordeboeke plaas van tweetalige woordeboeke tydens taalverwerwing of tydens die resepsieproses in vertaling (sien par. 5.1) gebruik word.

Kodewisseling is die verskynsel van gemengde taalgebruik by tweetalige persone. Waar taalverwarring egter onwillekeurig en skynbaar onbewus voorkom, is dit met kodewisseling nie die geval nie. Volgens Presas (2000: 26-27) is kodewisseling 'n meganisme wat die oordraghandeling onderlê. Die professionele vertaler gebruik die twee tale wat betrokke is *bewustelik* en *afwisselend*, maar wend verskillende vaardighede aan: "he or she receives in one language and produces in another". Presas meen die verwerwing van vertalingsbevoegdheid berus op 'n heroriëntasie van taalvermoë: van tweetalige bedrewenheid vorder die bekwame en ervare vertaler na tussentalige bedrewenheid.

3.1.2 Teksvermoë

Die professionele vertaler se linguïstiese besluite is ten nouste met die kommunikatiewe funksie van 'n teks verweef. Tekste vervul 'n bepaalde funksie in 'n kommunikatiewe situasie en hul kommunikatiewe sukses hang van die geslaagdheid van hul tekstuele samestelling en samehang af. Schäffner (2000: 146) omskryf tekstuele vermoë as kennis van die reëlmatigheid en konvensies van tekste, genres en tekstipes. Neubert (2000: 8) verduidelik dat woorde en (linguïstiese) strukture na gelang van die tekswêreld waarvan hulle deel is, patrone vorm en dat vertalers die normatiewe gebruik van sisteemelemente op verskillende terreine van diskoers probeer internaliseer. Naas die linguïstiese eise van 'n vertaalopdrag let hulle ook op die eise wat tekstuele eienskappe stel.

Die professionele vertaler wat vir die eerste keer met 'n bepaalde tekstipe of genre te doen kry, sal parallelle tekste² raadpleeg om reëlmatige eienskappe en konvensies te identifiseer en na te volg. In akademiese tekste word ander outeurs se menings en bevindinge byvoorbeeld aangehaal en word konvensionele uitdrukkings gebruik soos (*the author*) *relates (this) to (that), distinguishes between, argues that, looks at, illustrates, considers, outlines, highlights, offers, presents* en so meer. Die leksikograaf wat 'n woordeboek vir vertalers wil opstel, behoort korpusse van verskillende soorte tekste en genres (akademiese tekste, getuigskrifte, kleinadvertensies, weerberigte, resepte, gebruikshandleidings, ens.) te ontleed ten einde te bepaal watter reëlmatighede en konvensies daarin voorkom wat leksikografies bewerk en in 'n vertaalwoordeboek opgeneem kan word.

3.1.3 Vakvermoë

Professionele vertalers se ervaring maak dit vir hulle moontlik om die rol van vakspecialiste te speel — hulle stel hulle in die plek van dokters, regslui, aptekers of versekeringsagente en beheer die vaktaal asof hulle self in die praktyk staan (Robinson 1997: 148). Sommige vertalers spesialiseer in 'n bepaalde vakrigting, sodat hul vakkennis en beheer oor vakterme naderhand baie na is aan dié van die werklike vakkundiges op daardie terrein. Nogtans hoef hulle nie self alles te weet om steeds 'n goeie vertaling te kan voortbring nie, solank hulle weet waar om te gaan soek.

Die navrae wat praktiserende vertalers daagliks aan kollegas rig deur middel van ZaLang, 'n elektroniese gespreksgroep vir taalpraktisyns in Suid-Afrika, handel merendeels oor terminologie. Robinson (1997: 146) bevestig dat dit een van die belangrikste aspekte van 'n vertaler se taak is om gespesialiseerde woordeskat te hanteer. Kennis op alle terreine neem egter so vinnig toe dat 'n gedrukte woordeboek reeds onvolledig is wanneer dit verskyn. Professionele vertalers bou hul eie woordelyste van spesialisterme op, maak kontak-besonderhede bymekaar van vakkundiges en organisasies waar hulle hulp kan vra en weet hoe om parallelle tekste te raadpleeg ten einde styl, register en terminologiegebruik na te gaan.

Die moontlikheid van leksikografiese steun in dié verband verdien oorweging. 'n Vertaalwoordeboek kan moontlik die name van vakkundige verenigings en belangrike adresse of inligting oor elektroniese hulpbronne in die agterteksteafdeling insluit.

3.1.4 Kultuurvermoë

Engelse brontekste kan hul oorsprong in verskeie Engelssprekende lande hê — by die gebruik van korpusse vir lemmas moet die leksikograaf dit in ag neem. Nie alleen Suid-Afrikaanse Engelse tekste nie, maar byvoorbeeld ook Britse en Amerikaanse tekste moet gebruik word. Waar nodig moet geografiese etikette die sfeer van gebruik en moontlike betekenisverskille in die gebruik van 'n woord aandui. In Britse Engels kan 'n *instructor* byvoorbeeld omskryf word as "someone who teaches a skill such as driving or skiing", maar in Amerikaanse Engels word die woord ook gebruik om na "a schoolteacher or to a low-ranking university teacher" te verwys (Sinclair et al. 1997, s.v. *instructor*). Voorts kan die verskil tussen Britse, Amerikaanse en metrieke mate en gewigte, of tussen grade Fahrenheit en Celcius, met behulp van omrekeningstabelle in die agterteksteafdeling gegee word.

3.1.5 Probleemoplossingsvermoë

Professionele vertalers volg bepaalde strategieë ten einde vertaalprobleme op te los — sien in dié verband byvoorbeeld Chesterman (2000: 82), Newmark

(1988: 81-91), Shreve et al. (1993: 25), en Vinay en Darbelnet (1958/1995: 84-93). Probleemoplossing en kreatiewe denke loop hand aan hand. Kreatiewe denke behels onder meer die vermoë om vrylik soveel potensiële oplossings moontlik te genereer, hulle daarna te evalueer en die beste te kies. 'n Vertaalwoordeboek moet só saamgestel word dat dit die kreatiwiteit van die vertaler stimuleer en nie beperk nie.

3.1.6 Oordragvermoë

Vertalers se oordragvermoë sluit al die tegnieke, vaardighede en strategieë in wat hulle aanwend om 'n brontaaltekst so vinnig en doeltreffend moontlik in 'n funksionele doeltaaltekst oor te sit wat aan die vereistes van die vertaalopdrag sal voldoen. Dit betrek en integreer al hul ander vermoëns om tot 'n keuse tussen alternatiewe segswyses te lei. Soos Neubert (2000: 11) dit stel: "The actual transfer is the translator's decision. It results from selection."

4. Die algemene kenmerke van die vertaalproses

Om die spesifieke omstandighede te identifiseer waaronder 'n woordeboek aan professionele vertalers bystand kan verleen, is dit nodig om ook kortliks na die eienskappe van die vertaalproses te kyk.

Volgens Neubert en Shreve (1992: 43) behels die proses van vertaling "the activation of [translation] competence in a translation situation" — met ander woorde die aktivering van die vermoëns wat hierbo beskryf is. Die proses het die doeltelk as resultaat. Vertaling is teksproduksie. Tekste word geproduseer met die oogmerk om te kommunikeer en is dus bedoel om verstaan te word. Neubert en Shreve (1992: 46) noem produksie en resepsie die twee pole van die vertaalproses.

Die vertaalproses raak by ervare beroepslui taamlik outomaties en so te sê geroetineerd, maar selfs tegniese tekste wyk van algemeenhede en vaste patrone af en toets dan die kreatiwiteit van die vertaler. Daarom beskryf Kussmaul (1995: 40) die vertaalproses aan die hand van 'n kreatiwiteitsmodel, wat uit die volgende elemente of fases bestaan: (a) voorbereiding, (b) inkubasie, (c) verheldering en (d) evaluering. Dis belangrik om daarop te let dat hoewel die elemente hier onderskei word, hulle dikwels afwissel of ineenvloei en dit nie moontlik is om aan te dui waar die een fase ophou en 'n ander begin nie.³

In die voorbereidende fase ontleed die vertaler die brontekst en stel vas wat die doel van die doeltaal is. Moontlike begripsprobleme (resepsie) word opgelos deur kennis in te win en moontlike oplossings vir die probleme (produksie) word hier reeds bedink. Vir die tweede fase, inkubasie, is divergente denke nodig. Hoewel steeds deel van resepsie, is dit terselfdertyd ook reeds produksie. Die probleme is geïdentifiseer en oplossings word gegenereer: 'n klassieke kreatiewe proses. Divergente denke en 'n vlot gedagtevloei ("fluency") is noodsaaklik vir die kreatiewe denkproses en dus vir kreatiwiteit tydens

vertaling. In die protokolle wat Kussmaul (1995) beskryf, manifesteer 'n vlot gedagtevloei byvoorbeeld in die genereer van soveel moontlik alternatiewe segswyses ten einde die fase van verheldering te bereik, wanneer 'n moontlike oplossing hom voordoen. "The ability to find synonyms or semantically related words for a given expression seems to be a basic requirement for creative translation" (Kussmaul 1995: 51). Laastens evalueer die vertaler voortdurend moontlike oplossings ten einde vas te stel watter vir 'n bepaalde konteks toepaslik is.

In die beskrywing van haar eie werkswyse sê Bornemann (1989: 102) die sinonieme wat woordeboeke bied, wys vir vertalers dikwels bloot 'n algemene rigting aan: "Offered two synonyms, we will sometimes hit by association upon a third possibility, one which is not to be found in dictionaries yet, or at all, *in this context*" (beklemtoneering in die oorspronklike). Dis belangriker om die hele konteks indringend te bestudeer as om pasklaar vertaalekwivalente in woordeboeke te wil soek, soos onervare vertalers geneig is om te doen (Jensen en Jakobsen 2000: 113). Vir ervare vertalers stimuleer woordeboeke eerder die vertalingsproses: "The isolated word, even the isolated phrase — these nudge our thought processes" (Bornemann 1989: 99).

Dis belangrik dat 'n vertaalwoordeboek die vertaler se kreatiwiteit sal prikkel plaas van inhibeer — die pasklaar ekwivalente in tweetalige woordeboeke kan wel 'n inhiberende invloed op die vertaler se kreatiwiteit hê. Kussmaul (1995: 24) meen vertalers behoort eentalige verklarende woordeboeke plaas van tweetalige woordeboeke te gebruik wanneer hulle vertaal: "If we use [bilingual dictionaries] for translation purposes, the phase of abstraction, where we detach ourselves from the wording of a text, is completely suppressed, and the possibility of finding adequate translations for specific contexts by using our own imagination are very much reduced."

Woordeboekewivalente is nie noodwendig geskikte vertaalekwivalente nie, en dikwels is die vertaalekwivalent wat 'n kreatiewe vertaler bedink om 'n skeppende oplossing vir 'n vertaalprobleem te vind, nie gepas vir opname as woordeboekewivalent nie.

5. Die leksikografiese funksies en werklike doel van 'n vertaalwoordeboek

Soos hierbo aangedui, vorm die behoeftes van die gebruiker — trouens, die hele gebruikersprofiel — die grondslag vir die vasstelling van die leksikografiese funksies van 'n woordeboek. Die leksikografiese funksies van 'n vertaalwoordeboek verteenwoordig die hulp wat die woordeboek aan professionele vertalers bied ten einde hulle in staat te stel om probleme op te los wat tydens die vertaalproses voorkom. Die woordeboekstrukture en die keuse en aanbieding van leksikografiese gegewens word op die bereiking van die betrokke funksies gerig. Gesamentlik vorm al hierdie funksies en die veld wat die woordeboek dek, die werklike doel van die betrokke woordeboek (Bergenholtz en

Tarp 2003: 176). Op grond van Tarp en Gouws (2004: 2) se formulering van die werklike doel van 'n woordeboek, kan die werklike doel van 'n vertaalwoordeboek vir professionele vertalers van Engelse tekste in Afrikaans soos volg omskryf word:

Die werklike doel van 'n vertaalwoordeboek is om professionele vertalers wat die woordeboek raadpleeg ten einde 'n probleem op te los wat ontstaan terwyl hulle 'n teks uit Engels in Afrikaans vertaal, in staat te stel om die doel waarvoor hulle die woordeboek raadpleeg, te verwesenlik. Die werklike doel van 'n vertaalwoordeboek behoort dus te wees om 'n suksesvolle konsultasieproses in die hand te werk. Die sukses van die konsultasieproses hang af van die wyse waarop die vereiste linguïstiese inligting herwin kan word.

Na aanleiding van die beskrywing van die vertaalproses deur onder meer Neubert en Shreve (1992: 48), Mossop (1998: 40) en Kussmaul (1995: 40), asook op grond van Tarp (2002: 73), kan opsommend daarna verwys word as 'n proses waartydens die resepsie, oordrag en produksie van 'n teks plaasvind. In die vertaalproses wat hier bespreek word, is die resepsie, oordrag en produksie van 'n T2 > T1-tekste ter sake. Die professionele vertaler sal 'n Engels-Afrikaans-vertaalwoordeboek dus raadpleeg ten einde:

- 'n probleem op te los wat met die resepsie van die Engelse brontekste (T2) verband hou;
- 'n probleem op te los wat met die oordrag van die Engelse brontekste (T2) na 'n Afrikaanse doeltekste (T1) verband hou; en
- 'n probleem op te los wat met die produksie van die Afrikaanse doeltekste (T1) verband hou.

Die teorie van leksikografiese funksies maak voorsiening vir 'n tweedeling van woordeboekfunksies in kommunikasiegerigte en kennisgerigte funksies (Tarp en Gouws 2004: 4, en Bergenholtz en Tarp 2003: 173-174). In die drie gebruik-situasies wat hierbo genoem is, is die kommunikasiegerigte leksikografiese funksies van die beoogde vertaalwoordeboek dat dit professionele vertalers moet help met:

- die begrip van tekste in hul tweede taal;
- die vertaling van tekste uit hul tweede taal in hul eerste taal; en
- die produksie van tekste in hul eerste taal.

Die belangrikste kennisgerigte funksies van 'n vertaalwoordeboek is dat dit, waar nodig, aan professionele vertalers die volgende inligting moet verskaf:

- kulturele inligting;
- ensiklopediese inligting; en
- taalgebruiksinligting.

In die potensiële gebruikssituasie wanneer professionele vertalers tekste uit hul tweede taal (Engels) in hul eerste taal (Afrikaans) vertaal, is die kommunikasiegerigte funksies van oorheersende belang en word vervolgens veral aandag geskenk aan die wyse waarop 'n vertaalwoordeboek in hierdie omstandighede en aan hierdie gebruikersgroep bystand kan verleen.

5.1 Probleme met die resepsie van die T2-bronteks

The nature of language is such that, in the majority of cases, words have 'blurred edges'; their meanings are, to a large extent, negotiable and are only realized in specific contexts. ... One of the most difficult tasks that a translator is constantly faced with is that, notwithstanding the 'fuzziness' inherent in a language, s/he must attempt to perceive the meanings of words and utterances very precisely in order to render them in another language. (Baker 1992: 17)

Omdat konvensionele tweetalige woordeboeke gewoonlik daarop ingestel is om vertaalekwivalente te verskaf eerder as om die betekenis van woorde en uitdrukkings te verklaar, is hulle nie vir die oplos van begrips- of resepsieprobleme ideaal nie. Ofskoon Tarp (2002: 76) meen "mother-tongue equivalents would in most cases be sufficient and constitute the best solution", verhoed die naslaan van betekenis in 'n tweetalige woordeboek dat die T2-betekenis geïnternaliseer word as deel van 'n aparte repertorium gedagtebeelde of gedagteinhoud, soos in paragraaf 3.1.1 hierbo beskryf. Wanneer vertalers 'n eentalige verklarende woordeboek gebruik wat die betekenis van 'n woord, frase, uitdrukking of idioom in die T2-brontaal uiteensit, ondersteun en bevorder dit by hulle 'n gekoördineerde taalgeheue (par. 3.1.1). Dit spoor boonop hul skeppende verbeelding aan (par. 4) wat belangrik is, want oordrag en produksie is hier reeds aan die gang. Die implikasie hiervan is dat die betekenisverklaring in 'n vertaalwoordeboek in die brontaal moet geskied.

Eentalige verklarende woordeboeke maak dikwels op sinonieme staat om betekenis oor te dra. Dit illustreer 'n neiging om 'n leksikografiese proses te volg "whereby discrete areas of meaning are isolated from language in action" (Baker en Kaplan 1994: 3). Omdat vertalers egter met kommunikatiewe werklikhede op die vlak van diskoers eerder as met abstraksies werk, behoort die leksikografiese hantering van betekenisverklaring hierdie realiteit te erken en te weerspieël. Die implikasie hiervan is dat die betekenisverklaring van elke betekenisonderskeiding van die lemma deur middel van 'n kort omskrywing of parafrase moet geskied.

Dat die betekenisverklaring deur middel van definisies of parafrases gegee word, beteken nie dat tradisionele vorme van betekenisverklaring soos sinonieme, hiponieme en antonieme nie 'n plek het nie. Intendeel, in die lig van professionele vertalers se kreatiewe vermoë is dit goed moontlik dat juis een van dié brontaalvorme hulle in staat sal stel om 'n oplossing vir 'n vertaalprobleem te vind. Hoewel dit 'n fyn beplande mikroargitektuur sal vereis, is

die implikasie hiervan vir 'n vertaalwoordeboek dat veral brontaalsinonieme — en waar toepaslik brontaalantonieme — gelys moet word.

Die gebruik van werklike tekste waarop die inligting in 'n woordeboek gegrond word, bied volgens die outeurs van COBUILD (Sinclair et al. 1991: xv) die volgende moontlikhede: "It is possible to be precise about the shape of phrases and the extent of their variation; the relative importance of different senses of a word; and the typical environment in which a word or phrase is used". Die implikasie hiervan vir 'n vertaalwoordeboek is dat (a) sowel die lemmas as die vertaalekwivalente wat in die vertaalwoordeboek opgeneem word, op 'n korpus van bron- en doeltaaltekste gegrond moet wees wat werklike taalgebruik weerspieël en (b) die ordening van die betekenisonderskeidings die gebruiksfrekwensie daarvan moet weergee.

Dis nie net enkelwoorde se betekenis waarmee die vertaler begripsprobleme kan ervaar nie, maar ook multileksikale leksikale items soos frases, idioom, spreekwoorde, vreemdetaaluitdrukkings (bv. *sine qua non* of *faux pas*) en dies meer. Omdat professionele vertalers onder tyddruk werk, is dit belangrik dat die toegangstruktuur van 'n woordeboek hulle in staat sal stel om 'n frase of uitdrukking so vinnig moontlik na te slaan. Die implikasie hiervan vir 'n vertaalwoordeboek is dat (a) die tipografie en uitleg nie slegs ruimtebesparend moet wees nie, maar naslaanspoed moet bevorder en (b) meerwoordige items waar nodig hooflemmastatus moet kry.

5.2 Probleme met die oordrag van die T2-bronteks na die T1-doeltekst

In paragraaf 3.1.6 hierbo is aangedui dat die oordrag van 'n teks daaruit bestaan dat vertalers tussen alternatiewe segswyses 'n keuse maak. Om sodanige vertaalopsie te kan uitoefen, moet hulle óf self alternatiewe genereer óf gepaste vertalings elders vind, en dan besluit watter opsie die geskikste is. Die probleme wat vertalers hier kan noop om hul vertaalwoordeboek te raadpleeg, is met ander woorde:

- 'n Gebrek aan alternatiewe segswyses of vertaalekwivalente waaruit gekies kan word.
- Bevestiging dat die opsie wat gekies word, geskik is.

Die betekenis van die woord of uitdrukking in die brontaal is miskien vir die vertaler bekend, maar hy/sy kan nie dadelik aan 'n geskikte vertaalopsie dink nie. 'n Belangrike funksie van die vertaalwoordeboek in dié geval is om vertalers in staat te stel om vertaalopsies te genereer. Dit word gedoen deurdat die betekenisverklaring, voorbeeldmateriaal en sinonieme in die brontaal gegee word ten einde die vertaler se kreatiewe vermoë aan te wakker (sien par. 4). Indien die vertaler nie self op grond van hierdie prikkel 'n geskikte opsie kan genereer nie, is 'n volgende funksie van die vertaalwoordeboek om moontlike vertaalekwivalente van die lemma te verskaf.

5.3 Probleme met die produksie van die T1-doeltekst

Die produksie van 'n doeltekst bestaan daaruit dat 'n "correct and adequate text [...] in correspondence with the text in the source language" (Tarp 2002: 75) geskep moet word. Deur vertaalopsies aan die hand te doen, kan die vertaalwoordeboek aan die vertaler bevestig dat 'n moontlike vertaling wat die vertaler self vooraf geïdentifiseer het, wel 'n gepaste vertaalopsie is.

Voorts kan 'n vertaalwoordeboek vertalers van hulp wees met die oplos van ortografiese, sintaktiese, grammatikale of pragmatiese probleme op die vlak van die woord, kollokasie, idioom of spreekwoord (Tarp 2002: 75). By die produksie van T1-tekste is die funksie van die vertaalwoordeboek dus om aan die moedertaalspreker bystand te verleen deur middel van leksikografiese items wat bogenoemde aspekte aandui. 'n Belangrike implikasie vir 'n Engels-Afrikaans-vertaalwoordeboek is dat die lemma waaronder al hierdie items geplaas word, 'n doeltaallemma behoort te wees ten einde die vertaler se soektog te vergemaklik. Op grond van hierdie oorweging is 'n tweerigtingwoordeboek waarskynlik gewens. Vir die doel van hierdie artikel word egter op die ontwikkeling van 'n eenrigtingwoordeboek gekonsentreer.

6. Die begrippe tweetalige, vertalende en vertaalwoordeboek

Uit die beskrywing hierbo van 'n woordeboek wat op die behoeftes van professionele vertalers toegespits is, is dit duidelik dat sodanige woordeboek anders sal lyk as wat die meeste tweetalige woordeboeke tradisioneel daar uitsien. Dis dus gepas om in dié stadium ondersoek in te stel na wat met die begrippe *tweetalige woordeboek*, *vertalende woordeboek* en *vertaalwoordeboek* bedoel word.

6.1 Tweetalige woordeboek

Wanneer vertalers na hul "tweetalige woordeboek" verwys (teenoor hul "verklarende woordeboek", waarmee hulle 'n eentalige woordeboek soos die *Handwoordeboek van die Afrikaanse Taal* (HAT) bedoel waarin die leksikale betekenis van lemmas verklaar word), het hulle gewoonlik 'n woordeboek soos die *Tweetalige Woordeboek/Bilingual Dictionary* in gedagte: 'n woordeboek wat uit twee helftes bestaan waarin die twee tale wat die behandelde taalpaar vorm, om die beurt eers as brontaal en dan as doeltaal optree. Vertaalekwivalente in die doeltaal word vir elke brontaallemma aangebied en die twee kante vorm naastebly 'n spieëlbeeld van mekaar.

Woordeboeke wat vertaalekwivalente aanbied, staan in die metaleksikografie onder verskillende name bekend. Hulle word nie slegs tweetalige woordeboeke ("bilingual dictionaries") genoem nie, maar ook tussentalige woordeboeke ("interlingual dictionaries"), vertalende woordeboeke ("translation dictionaries" of "translatory dictionaries"), of tweetalige vertalende woordeboeke.

Die terme word dikwels uitruilbaar gebruik. Burkhanov (2004: 4) meen dat daar "[a] good deal of inconsistency in metalexigraphic considerations concerning bilingual lexicography" is en dat dit onder meer toegeskryf kan word aan leksikograwe se verskillende vertolking van wat "vertaling" behels.

Snell-Hornby (1984: 275) meen tweetalige leksikografie behoort beter gedifferensieer te word, aangesien soveel verskillende teikengroepe ter sprake is. Vir toeriste bly 'n sakwoordeboek waardevol, vir tegnici bied 'n tweetalige woordeboek met tegniese terme en gestandaardiseerde konsepte die ideale oplossing, en vir die aanleerders van 'n vreemde taal is spesiaal saamgestelde tweetalige woordeboeke onontbeerlik.

In die metaleksikografie geniet die verfyning van terminologie en tipologiese klassifikasie deurlopend aandag. Hausmann en Werner (1989–1991: 2740–2743) bied byvoorbeeld 'n uitgebreide bespreking van tweetalige woordeboeke op grond van taalrigting of "skopus" (eenrigting of tweerigting), direksionaliteit (afhangend van die woordeboek of woordeboekdeel se gerigtheid op die gebruiker as moedertaal- of tweedetaalspreker) en doel of funksie van die woordeboek (aanleer van 'n taal of hulp met probleme by die omgaan met tekste — by laasgenoemde word ook tussen aktiewe en passiewe funksies onderskei, na gelang daarvan of resepsie of produksie ter sprake is).

Tarp (2002: 60–63) toon egter dat 'n algemeen aanvaarde definisie van 'n tweetalige woordeboek nie bestaan nie. In hierdie artikel word die begrip in sy gewone sin gebruik.

6.2 Vertalende woordeboek

In 1987 vind 'n kollokwium met die tema "Vertaling en Leksikografie" aan die Universiteit van Innsbruck plaas met die medewerking van die Instituut vir Vertaling en Tolkwerk. Die rede waarom spesialiste in vertaling en leksikografie byeenkom, is eerstens die gebrek aan geskikte woordeboeke vir professionele vertalers en vertalersopleiding, en tweedens die skynbare onbewustheid van woordeboekmakers van die spesiale behoeftes van vertalers — 'n belangrike groep gebruikers van leksikografiese produkte (Snell-Hornby, Pöhl en Bennani 1989: 5).

Op die kollokwium maak Hartmann (1989: 9) die stelling dat "the bilingual dictionary is often called the translation dictionary par excellence because it offers the user interlingual lexical equivalents".

In die Afrikaanse metaleksikografie word na die konsep "translation dictionary" as 'n vertalende woordeboek verwys. Die gebruik van dié term impliseer dat die woordeboek as die aktiewe agent by vertaling beskou word. By verdere implikasie kan 'n maatskappysekretaresse byvoorbeeld met behulp van 'n vertalende woordeboek haar maatskappy se dokumente vertaal — 'n suggestie wat versterk word as leksikograwe meen 'n vertalende woordeboek mag nie op die intuïsie van die gebruiker staatmaak om die regte vertaalekwivalent te vind nie (Gouws 1992: 38).

Die keuse van korrekte ekwivalente impliseer egter nie noodwendig 'n goeie vertaling nie. Bestaande tweetalige vertalende woordeboeke is geskik vir die deursneegebruiker wat na vertaalekwivalente soek, maar voldoen nie optimaal aan die behoeftes van professionele vertalers nie.

6.3 Vertaalwoordeboek

Die professionele vertaler benodig leksikografiese ondersteuning met die oplos van probleme wat by die vertaling van 'n bronteks in 'n doelteks kan voorkom, met ander woorde tydens die resepsie, oordrag en produksie van 'n teks. In die lig van die bespreking hierbo van wat dié proses behels, is die term "vertalende woordeboek" onvanpas vir 'n woordeboek wat op die behoeftes van professionele vertalers gerig is.

Professionele vertalers beskou die doeltaalwoorde en -uitdrukkings wat die leksikograaf voorsien nie as *die* vertaling van brontaalitems nie, maar as "translated cognitive orientations" — 'n punt vanwaar hulle self na 'n gepaste vertaling kan soek: "The term translation dictionary does not entail that it offers the needed translation. It points the way to the translation, no more" (Neubert 1992: 34-35). Die term "vertaalwoordeboek" is in Afrikaans 'n toepasliker benaming vir 'n woordeboek wat daarop gerig is om dié soort ondersteuning te bied.

7. Ten slotte

Uit die voorafgaande bespreking is dit duidelik dat 'n woordeboek wat spesifiek op professionele vertalers se behoeftes gerig wil wees, op 'n vernuwende wyse bewerk en aangebied moet word.

Notas

1. Die konsep van volkome tweetaligheid is relatief. Tweetaligheid is nie 'n punt waarby iemand finaal arriveer nie, maar 'n kontinuum. Dis nie moontlik om 'n vlak van tweedetaalvaardigheid te identifiseer as die vlak van beheer waarop iemand as ten volle tweetalig gereken kan word nie.
2. Neubert (2000: 151) beskryf *parallele tekste* as "L2 and L1 texts of equal informativity which have been produced in more or less identical communicative situations".
3. Neubert en Shreve (1992: 50) bevestig dat die vertaalproses nie as 'n eenvoudige, opeenvolgende reeks prosedures gesien moet word wat in 'n vaste volgorde geaktiveer word nie: "The entire convoluted process resembles a network more than a chain. [...] The progression of text comprehension and text production procedures are probably guided by problem-solving procedures, assisted by information recall, storage, and integration procedures, and monitored by pattern-matching and planning procedures."

Bronnelys

Woordeboeke

- Bosman, D.B., I.W. van der Merwe en L.W. Hiemstra (Reds.).** 1984⁸. *Tweetalige Woordeboek/Bilingual Dictionary*. Kaapstad: Tafelberg-Uitgewers.
- Odendal, F.F. en R.H. Gouws (Reds.).** 2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT). Kaapstad: Pearson Education South Africa.
- Sinclair, J., G. Fox en S. Bullon (Reds.).** 1997. *Collins COBUILD English Dictionary*. Londen: HarperCollins.

Ander

- Baker, Mona.** 1992. *In Other Words: A Coursebook on Translation*. Londen/New York: Routledge.
- Baker, Mona en Robert Kaplan.** 1994. Translated! A New Breed of Bilingual Dictionaries. *Babel* 40(1): 1-11.
- Bergenholtz, Henning en Sven Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31-2003: 171-196.
- Bornemann, Eva.** 1989. Translation and Lexicography: A Practical View. Snell-Hornby, M., E. Pöhl en B. Bennani (Reds.). *Translation and Lexicography: Papers Read at the EURALEX Colloquium held at Innsbruck 2-5 July 1987*: 99-104. Philadelphia: John Benjamins.
- Burkhanov, Igor.** 2004. Requirements for an "Ideal" Bilingual L1 > L2 Translation-oriented Dictionary. *Lexikos* 14: 1-19.
- Chesterman, Andrew.** 2000. Strategies for Emancipatory Translation. Schäffner, C. en B. Adab (Reds.). *Developing Translation Competence*: 77-89. Amsterdam/Philadelphia: John Benjamins.
- Gouws, Rufus H.** 1992. Kommunikatiewe ekwivalensie in vertalende woordeboeke — 'n probleem vir die vertaler. *Suid-Afrikaanse Tydskrif vir Taalkunde* 10(1): 37-42.
- Hartmann, Reinhard R.K.** 1989. Lexicography, Translation and the So-called Language Barrier. Snell-Hornby, M., E. Pöhl en B. Bennani (Reds.). *Translation and Lexicography. Papers Read at the EURALEX Colloquium Held at Innsbruck 2-5 July 1987*: 9-20. Philadelphia: John Benjamins.
- Hausmann, Franz Josef en Reinhold Otto Werner.** 1989-1991. Spezifische Bauteile und Strukturen zweisprachiger Wörterbücher: Eine Übersicht. Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopaedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 2729-2778. Berlyn/New York: De Gruyter.
- Jensen, Astrid en Arnt Lykke Jakobsen.** 2000. Translating under Time Pressure: An Investigation of Problem-solving Activity and Translation Strategies by Non-professional and Professional Translators. Chesterman, A., N. Gallardo San Salvador en Y. Gambier (Reds.). *Translation in Context: Selected Contributions from the EST Congress, Granada 1998*: 105-116. Amsterdam: John Benjamins.
- Kussmaul, Paul.** 1995. *Training the Translator*. Amsterdam/Philadelphia: John Benjamins.
- Mossop, Brian.** 1998. The Workplace Procedures of Professional Translators. Chesterman, A., N. Gallardo San Salvador en Y. Gambier (Reds.). *Translation in Context: Selected Contributions from the EST Congress, Granada 1998*: 39-48. Amsterdam: John Benjamins.

- Neubert, Albrecht.** 1992. Fact and Fiction of the Bilingual Dictionary. *EURALEX '90 Proceedings. IV International Congress*: 29-42. Barcelona: Bibliograf.
- Neubert, Albrecht.** 2000. Competence in Language, in Languages, and in Translation. Schäffner, C. en B. Adab (Reds.). *Developing Translation Competence*: 3-18. Amsterdam/Philadelphia: John Benjamins.
- Neubert, Albrecht en Gregory M. Shreve.** 1992. *Translation as Text*. Kent, Ohio: Kent State University.
- Newmark, Peter.** 1988. *A Textbook of Translation*. Londen: Prentice-Hall.
- Presas, Marisa.** 2000. Bilingual Competence and Translation Competence. Schäffner, C. en B. Adab (Reds.). *Developing Translation Competence*: 19-31. Amsterdam/Philadelphia: John Benjamins.
- Roberts, Roda P.** 1990. Translation and the Bilingual Dictionary. *Meta* 35(1): 74-82.
- Robinson, Douglas.** 1997. *Becoming a Translator: An Accelerated Course*. Londen/New York: Routledge.
- Schäffner, Christina.** 2000. Running before Walking? Designing a Translation Programme at Undergraduate Level. Schäffner, C. en B. Adab (Reds.). *Developing Translation Competence*: 144-156. Amsterdam/Philadelphia: John Benjamins.
- Shreve, G.M., C. Schäffner, J.H. Danks en J Griffin.** 1993. Is there a Special Kind of "Reading" for Translation? An Empirical Investigation of Reading in the Translation Process. *Target* 5(1): 21-41.
- Snell-Hornby, Mary.** 1984. The Bilingual Dictionary: Help or Hindrance? Hartmann, R.R.K. (Red.). *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983*: 274-281. Tübingen: Max Niemeyer.
- Snell-Hornby, Mary en Esther Pöhl.** 1989. Preface. Snell-Hornby, M., E. Pöhl en B. Bennani (Reds.). 1989. *Translation and Lexicography: Papers Read at the EURALEX Colloquium Held at Innsbruck 2-5 July 1987*. Philadelphia: John Benjamins.
- Tarp, Sven.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, Sven.** 2002. Translation Dictionaries and Bilingual Dictionaries — Two Different Concepts. *Journal of Translation Studies* 7: 59-84.
- Tarp, Sven en Rufus H. Gouws.** 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke? *Tydskrif vir Geesteswetenskappe* 44(4): 276-298.
- Vinay, Jean-Paul en Jean Darbelnet.** 1958/1995. *Comparative Stylistics of French and English: A Methodology for Translation*. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, Herbert Ernst.** 1988. Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis van sprachlichem und enzyklopädischem Wissen. Munske, H., et al. (Reds.). 1988. *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern*: 729-790. Berlyn/New York: Walter de Gruyter.

Compiling a Corpus-based Dictionary Grammar: An Example for Northern Sotho

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),*
and

Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (elsabe.taljard@up.ac.za)*

Abstract: In this article it is shown how a corpus-based dictionary grammar may be compiled — that is, a mini-grammar fully based on corpus data and specifically written for use in and integrated with a dictionary. Such an effort is, to the best of our knowledge, a world's first. We exemplify our approach for a Northern Sotho mini-grammar, to be included into a Northern Sotho–English dictionary.

Keywords: LEXICOGRAPHY, DICTIONARY, CORPUS, FREQUENCY, MIDDLE MATTER, DICTIONARY GRAMMAR, NORTHERN SOTHO (SESOTHO SA LEOA)

Samenvatting: Een corpusgebaseerde woordenboekgrammatica samenstellen: een voorbeeld voor Noord-Sotho. In dit artikel word aangetoon hoe een corpusgebaseerde woordenboekgrammatica kan saamgestel word — dit is, een minigrammatica die al z'n gegevens regstreeks uit een corpus haal en die spesiaal geskryf word om in een woordeboek gebruik te word, en er ook volledig mee geïntegreer is. Zo'n poging is, voor zover ons bekend, een wereldprimeur. We illustreer ons aanpak voor een minigrammatica van het Noord-Sotho, bedoel om gebruik te word in een Noord-Sotho–Engels woordeboek.

Sleutelwoorde: LEXICOGRAFIE, WOORDENBOEK, CORPUS, FREQUENTIE, MIDDENWERK, WOORDENBOEKGRAMMATICA, NOORD-SOTHO

1. Using corpora beyond a dictionary's central section(s)

It is now widely accepted that the use of electronic corpora has become indispensable in modern dictionary making, and this on a variety of levels. But just on how many levels? The macrostructural and microstructural levels immediately spring to mind, and most attention in the scientific literature has indeed also gone to aspects revolving around the corpus-based selection of lemma signs on the one hand, and the corpus-based construction of articles attached to

those lemma signs on the other. Any self-respecting dictionary, however, contains much more than 'just' the central text. Good dictionaries also comprise extra matter, invariably distributed across front, middle and back matter sections. If one is serious about corpus-based lexicography, then the extra matter should also be rooted in corpus data. One can come a long way by making sure there is a one-to-one correlation between the central (corpus-based) section(s) and the extra matter (cf. below), but during practical dictionary making this quickly proves not to be sufficient.

In this article the focus will be on the creation of a corpus-based dictionary grammar, exemplified for Northern Sotho. The core principles of corpus-based lexicography will be briefly reviewed in order to set the stage, but that review is merely incidental and the reader is referred to Sinclair (1987) and Corr  ard (2002) for what remain to this day the best collections on the topic.

2. Corpus-based lexicography in a nutshell

In corpus-based lexicography, the main arbiter during the creation of the (initial) macrostructure is the list of frequencies attached to the lemmatised list of inclusion candidates. Clearly, there are as many lemmatisation policies as there are dictionary teams compiling dictionaries, but it remains common practice to work with and label homonyms, part-of-speech groups, senses (often structured as a hierarchy), and translation equivalents. On each of these levels, frequency considerations are used to order the items, with the more frequent ones presented before the lesser frequent items. In bidirectional explanatory dictionaries, each side draws its data from its own dedicated corpus, so there one has to do this entire exercise twice. Moving to the hard part of the microstructural level, the core process is to derive meanings from the uses as seen in the corpora queried. Each of the uses is then exemplified with corpus extracts — the very extracts that led the lexicographers to divide their articles into senses in the first place. Combinations, derivations, idiomatic expressions, and the like, are all selected by considering their respective occurrence frequencies. Misspellings, misuses (exploitations of the norms?), and anything striking is also noted and finds its way into the dictionary; for example by cross-referring erroneous forms to the correct ones, or by means of the inclusion of usage notes. In short, the very structure of each and every article, even though conforming to a strict and well-defined DTD (document type definition), is 'inspired' by what the dictionary compiler sees in and distils from the corpus. Using a corpus is not a once-off process, but rather a continuous one. This process is greatly facilitated by software where the dictionary writing system (DWS) and the corpus query package (CQP) are seamlessly integrated (De Schryver and De Pauw 2007).

As noted in Section 1, in addition to a dictionary's macrostructure(s) and microstructure(s), most dictionaries also contain extra matter material. Numerous examples exist of dictionaries where the extra matter was clearly written in

isolation, and if not to that extreme, at least not with a full integration in mind. It is classic that front matter descriptions on 'How to Use Your Dictionary' contain examples that, when followed up, lead nowhere — as the very examples used to exemplify the usage, have not been included in the dictionary. Even a cursory glance through dictionary reviews reveals this. In addition, for languages where the orthography has not yet been standardised, one needs to be extra careful to spell the material in the extra matter section(s) in the same way as was done in the central section(s) — and of course one should adhere to consistency within all of these sections as well. Here, too, recourse to a corpus during the compilation seems wise.

In a recent dictionary project, a particularly exciting aspect was the attempt to compile a mini-grammar, based, just as all other aspects of that dictionary, on corpus data. This project, as well as the core features for the write-up of a corpus-based mini-grammar, is discussed in the next section.

3. Isolating the core features for a corpus-based mini-grammar

Given the project is still under embargo, we can unfortunately not reveal the Publisher, nor can we give too many (trade) details. Future publications in the scientific literature will no doubt deal with the larger metalexicographical picture of this project. What can be revealed at this stage are the following features:

- Bidirectional, bilingual Northern Sotho–English dictionary
- Aimed at a well-defined junior target user group, but also with more advanced users in mind (given no other dictionaries are available for this market)
- Corpus-based mini-grammar is part of the middle matter
- Very few pages only for the mini-grammar, eight in all, and this for both a Northern Sotho and an English version (thus four pages each)

From this it is clear that one must attempt to (a) describe the absolute core, and (b) make sure the mini-grammar truly fits and blends into the other sections of the dictionary. The second issue is the easiest, as it means one simply has to attempt to be as consistent as possible at all times: the same abbreviations and conventions as those used in the central sections must be employed, the meta-language must be presented in the same language as seen in the central sections, all examples given must also be in the dictionary itself, etc. When it comes to the first issue, it seems one could merely attempt to isolate, say, the ten most important grammatical features of Northern Sotho, and to present those. Indeed, and as will be seen in Section 4, some grammar aspects are frequent, while others are not, so theoretically one could just select the top ten based on frequency considerations.

Book-length, fully corpus-based grammars have been written in this way, especially for English. The best known of these is the *Longman Grammar of Spo-*

ken and Written English (Biber et al. 1999). Unlike the latter, for a corpus-based mini-grammar one does not have hundreds and hundreds of pages to describe minute corpus-based phenomena, and also unlike the latter, a corpus-based mini-grammar is not a generic stand-alone description, it is intended to be an integral part of a very specific product. In our case this means that one must keep both the product and its user in mind. Firstly, the product is a bilingual Northern Sotho–English dictionary, which means that one must also cover, even 'bridge', Northern Sotho and English. Reformulated, English grammar, if not the focus of the description, does play a role, and must be contrasted with Northern Sotho where necessary, and in some cases even take centre stage. Secondly, given the target user is not advanced, initially, it means that complicated grammatical issues must be described in as easy a language as possible, but without sacrificing too much accuracy. (After all, specialised vocabulary does have a *raison d'être*.) To the best of our knowledge, the compilation of a corpus-based dictionary grammar, no matter its size, has never been attempted before.¹

Terra cognita. It is well-known that lexicographers struggle to find 'the right place' for anything that is not part of their dictionary's central section(s). Users simply do not seem to find their way to any extra matter, and in recent learners' dictionaries, compilers go as far as randomly interspersing their dictionary text with full colour plates or entire reference sections. No studies have reported on the efficacy of this approach, however. Nonetheless, in bilingual lexicography, there are three 'logical' places to present extra matter: preceding the first side of the dictionary (i.e. the front matter), between the two sides (i.e. the middle matter), and following the second side (i.e. the back matter). In our case we chose to make the mini-grammar part of the middle matter, which also includes other material that belongs to a so-called 'Study section'. The publisher is responsible for making this section stand out visually from the core dictionary sections, yet not to such an extent that it is easily skipped. The eight pages were divided as follows: three pages of text in Northern Sotho and three pages of text in English, with between these, two pages for tables and figures with double labelling so they fit both Northern Sotho and English. This seemed to be the most economical use of the allowed space.

Terra incognita. We decided to single out ten topics for the mini-grammar. Following frequency checks of both Northern Sotho as well as English topics, the decision was made to devote seven frequent topics to Northern Sotho, two frequent topics to English, and one *non*-frequent topic to Northern Sotho. The group of seven seems straightforward, and is also what intuition would have dictated. Given the bilingual imbedding of the material, the group of two, which deals with important (English) issues that are often a cause for confusion, can also be explained. The singleton group, which deals with an issue that is not even covered in the dictionary, will clearly need further explanation (cf. below, Section 4.10). The short headings for each of the ten topics are as follows:

-
- Articles
 - Nouns
 - Nominal suffixes
 - Verbal suffixes, verbal prefixes, negative verbs
 - Agreement system
 - Adjectives
 - Pronouns
 - Demonstratives
 - Locative particles
 - Tone

The sections on 'Articles' and 'Pronouns' deal with English grammar, while the section on 'Tone' has no counterpart in the central section of the dictionary. With regard to the order of the ten topics, frequency considerations were again considered. What is not covered is presented last. Perhaps surprisingly, the issue that leads the list is basically an English issue, as it deals with the English definite and indefinite articles (and we all know that "the" leads any English-language lemmatised frequency list). The order Nouns > Verbs > Adjectives should also not surprise. In short, the ten topics were sorted from most important to lesser important, amalgamating both Northern Sotho and English frequencies. In the next section the mini-grammar itself will be presented, accompanied by liner notes for each.

4. Corpus-based mini-grammar for a Northern Sotho–English dictionary

4.0 Introduction

Northern Sotho belongs to a large family of languages, internationally known as the Bantu language family. Most of the languages spoken in the southern half of Africa belong to this group, and therefore share many grammatical features. Approximately 4.2 million people in the Republic of South Africa speak Northern Sotho as a home language. The main characteristic features of Northern Sotho are summarized in this section. Note that the tables and figures on pages X and Y accompany this summary.

The grammar is introduced by this paragraph, where, from the start, the dictionary user is alerted to the fact that various tables and figures accompany the text. In this introduction, we look both back (it is a Bantu language) and forward (spoken in South Africa, by 4.2 million people).

4.1 Articles

❶ There are *no definite or indefinite articles* in Northern Sotho. This means that there are no equivalents for the English words 'the', 'a' or 'an'.

Starting a mini-grammar of Northern Sotho with what is basically an English grammatical point may seem counterproductive, but was prompted by various

frequency considerations. Not only is "the" the top-frequent English lemma sign (with "a" in fifth position, and "an" in 33rd), real dictionary usage logs involving a Bantu language on the one hand and English on the other, confirm that both the definite article "the" and the indefinite article "a" are among the most frequently searched-for items (cf. De Schryver et al. 2006: 76). It is thus imperative to point out, right from the start, that these articles simply do not have corresponding translation equivalents in Northern Sotho. The lemma sign "a" of course also heads the first article of the English to Northern Sotho side of the dictionary. There, a usage note gives further information (and also focuses more on English grammar).

4.2 Nouns

☛ **Nouns** are grouped into classes, which are numbered according to an internationally accepted numbering system. The class to which a noun belongs can be identified by looking at the first part of the noun, which is called the noun class prefix (CP, cf. Table 1). Classes 1 to 10 are arranged **in pairs** with the unevenly numbered classes (1, 3, 5, 7, 9) containing singular forms, and the evenly numbered ones (2, 4, 6, 8, 10) the corresponding plural forms. In your dictionary, nouns are entered under their singular form, and plurals need to be looked up under their corresponding singular form. Also, the class of the headword is shown in bold, together with the corresponding one in non-bold.

Not all nouns have both singular and plural forms — some only ever occur in the singular form, whereas others only have a plural form. In your dictionary, the absence of either singular or plural form is indicated by a dash (–) in the slot where class membership is indicated, for example *setu* 'silence' shows 7/–. Nouns in class 14 normally do not have a plural form and those that do, use the plural prefix of class 6. Nouns in classes 16, 17, 18 and two unnumbered classes (sometimes called the *N*-class and the *ga*-class or class 24) refer to spatial orientation and are called **locative classes**. The locative classes, together with class 15 (the infinitive class), do not distinguish plural forms.

Nouns are by far the most frequent part of speech in the Northern Sotho to English side of the dictionary, as they account for 51% of all entries. Northern Sotho nouns are also the most important words in any sentence, in the sense that the way a sentence *looks* (cf. Section 4.5) is entirely dependent on the class prefix of the noun. Given that the composition of nouns (class prefix + noun stem) is entirely different from that of English nouns, and given that especially the way plurals are formed is very different (in Northern Sotho the prefix changes, compared to the addition of *-s* or *-es* at the end in English), a separate section was needed. (In contrast, basic Northern Sotho verbs work very much like verbs in English, so a separate section on the basic Northern Sotho verb was not needed.)

Also note that guidance is given (here, and throughout the mini-grammar) on how the grammatical characteristics impact on the way the dictionary was compiled, and thus on how certain words should be looked up.

The mentioned 'Table 1' as well as all other tables and figures from the mini-grammar have been reproduced at the end of this article.

4.3 Nominal suffixes

Ⓢ A number of **nominal suffixes** can be added to a noun to change its meaning. One such suffix is the **locative marker -ng**. Adding *-ng* to a noun adds the meaning 'in/at/to/from' to the meaning of the noun. For example: *toropo* 'town', but *toropong* 'in town'. In your dictionary, all frequent cases of nouns with locative markers are treated as derivations under the main noun. Another important nominal suffix is the **diminutive suffix -ana**, used to express 'small, little, short, etc.', such as in *mokotla* 'bag' versus *mokotlana* 'small bag'. Because of sound changes caused by this suffix, for example *kgarebe* 'young girl' versus *kgarebjana* 'little girl', such forms have been entered as headwords in your dictionary.

Based on frequency considerations, two types of words were lemmatised *with* their nominal suffixes in the dictionary. The first, the locative marker *-ng*, does not cause any (major) sound changes, so nouns with this suffix have simply been lemmatised as sub-lemmas. The second, the diminutive *-ana*, can cause more important sound changes, and with the target user group in mind, such nouns have been lemmatised. The function of this third topic in the mini-grammar is thus dual: firstly to point out that suffixes can be attached to nouns, and secondly to explain where to find nouns with suffixes in the dictionary.

Implicit in this exposition is of course that what are prepositions in English, can be simply nominal suffixes in Northern Sotho. This is but one of many mismatches in parts of speech between the treated language pair.

4.4 Verbal suffixes, verbal prefixes, negative verbs

Ⓢ **Suffixes** can also be added to **verbs**, such as the **relative marker -go (or -ng)**, the **plural marker -ng**, or what are known as the **verbal extensions**. The most frequently used single verbal extensions are listed in Table 3. The actual form taken on by these suffixes may vary, following certain phonological rules. Common combinations of verbal extensions are shown in Table 4. Figure 1 shows how frequent the most important verbal extensions are. In your dictionary, the presence of verbal extensions is always indicated, with a cross-reference to the verb stem (when present).

Verbal prefixes such as subject concords (SC) and object concords (OC) are usually separated from the verb stem, but three kinds of prefixes are fixed to the verb stem, namely the **reflexive prefix i-** 'self', the **OC of the first person singular n-/m-** 'me', and the **OC of class 1 m-** 'her/him'. For the latter, this happens only when the verb which follows starts with a *b-*. Examples: *ruta* 'teach' > *ithuta* 'learn; teach oneself', *thuša* 'help' > *nthuša* 'help me', *botša* 'tell' > *mmotša* 'tell her/him'. In your dictionary, all frequent verbs with attached prefixes have been entered.

The **negative forms** of verbs are formed by means of the **negative morphemes ga, sa and se**, which appear as verbal prefixes. In some cases, the use of these morphemes cause the verbal ending to change from *-a* to *-e*. For example: *Banna ba aga sekolo*. 'The men are building a school.' > *Banna ga ba age sekolo*. 'The men are not building a school.' In your dictionary, negative forms have been entered as combinations under verbs that end in *-e*, and are preceded by **'ga/sa/se (...)**.

This fourth topic of the mini-grammar deals with verbs, good for 37% of all entries in the dictionary. The focus is not on the verb stems (i.e. the basic verbs),

as these are straightforward, but on the various suffixes, prefixes and negative forms that surround verb stems.

The suffixes *-go* and *-ng* are non-problematic, which is why a mere mention is adequate. However, no less than 57% of all the verbs entered into the dictionary contain one or more verbal extensions, the orthographic form of which is governed by a multitude of complex rules, which is why Tables 3 and 4, as well as Figure 1, devote a considerable amount of space to this issue. The focus, once again, is on the frequent possibilities only. This is especially clear from Figure 1.

The second paragraph deals with those verbs where prefixes have been written conjunctively, which occurs for reflexive verbs (76 cases in the dictionary), object concords of the first person singular (31 cases), and object concords of class 1 for *b*-initial verbs (13 cases). These forms have been lemmatised as such, as the various morphophonological sound changes make it very hard to isolate the stems of these verbs.

In the third paragraph the so-called *ga/sa/se*-convention for Northern Sotho is briefly described. This convention was introduced by Prinsloo and Gouws (1996), and is a useful (approximate) tool to summarise numerous (negative) tenses into just one compact dictionary article. Negative morphemes can of course be used in combination with any type of verb, including those with pre- and suffixes. In all, 14% of all verbs entered into the dictionary are so-called 'negative verbs'.

4.5 Agreement system

⑥ Northern Sotho has a complex linking system (also called an **agreement system**) in which nouns are linked by means of concords to verbs, adjectives, pronouns and other nouns. Subjects for example, are linked to verbs by means of subject concords (SC): *Baithuti **ba** bala Seisimane*. 'The students read English.' When the subject is deleted, these concords function as pronouns: ***Ba** bala Seisimane*. 'They read English.' In the phrase *mmotoro **wa** gagwe* 'her car' (literally 'the car of her'), the possessive concord (PC) *wa* is the link between the possession *mmotoro* and the possessor *gagwe*. The form of the concord depends on the possessor noun, in this case *mmotoro*. If a noun from a different class is used, the concord will change, for example: *dipuku **tša** gagwe* 'her books'. See Tables 1 and 2 for all concords.

The noun class system together with the linked concordial agreement system forms the heart of the Northern Sotho grammar. Once nouns and noun classes (Sections 4.2 and 4.3) as well as verbs (Section 4.4) have been introduced, it can be presented. Due to the complexity, one has no other option than to do this in tabular form, accompanied by selected examples. What is rather unique in this corpus-based presentation, however, is that all frequent forms have been singled out and are typographically different from the lesser or non-frequent forms. Indeed, in Tables 1 and 2 all the frequent concords are printed in bold, and these also correspond with those — and only those — items that have been lemmatised in the dictionary.

4.6 Adjectives

⑥ In English, we say 'red dress', thus the adjective 'red' precedes the noun which it describes. In **Northern Sotho**, **adjectives** follow the noun: *roko ye khubedu*, where *roko* 'dress' is the noun, and *ye khubedu* 'red' the adjective. Northern Sotho adjectives are made up of an adjective stem (of which there are no more than 30), in most cases preceded by a corresponding class prefix (CP) fixed to this adjective stem, and also a preceding demonstrative (DEM). In both sides of your dictionary, all the frequent adjectives have been entered in full, with their corresponding demonstrative also being shown. See for example the entry for 'black' in your dictionary.

As a result of the limited number of adjective stems, other grammatical constructions are used to describe nouns in Northern Sotho. These constructions often correspond to **English adjectives**. They are the following:

- Possessive construction, in your dictionary '[PC +]', which consists of a possessive concord (PC) followed by a noun: *meetse a borutho* 'warm water' (literally 'water of warmth').
- Nominal relative construction, in your dictionary '[DEM +]', which consists of a demonstrative (DEM), followed by a relative noun. Relative nouns often belong to class 14: *bophelo bjo bonolo* 'easy life'.
- Verbal relative construction, in your dictionary '[DEM + SC +]', which consists of a demonstrative (DEM) plus a subject concord (SC), followed by a verb that has the relative suffix **-go**: *mamapo a a elago* 'liquid honey'.

In contrast to Northern Sotho, the English language has hundreds of adjectives (there are around 660 in your dictionary). On the English–Northern Sotho side, adjectives that are not frequent in Northern Sotho are therefore abbreviated. For instance, under 'enormous': **[DEM +] CP**gologolo; **[DEM +]** kgolokgolo. The form with 'CP' (the class prefix) is valid for all classes except 8 to 10, while the other form is valid for classes 8 to 10 only.

In the Bantu languages, there are only about 30 adjective stems. These stems take a class prefix, but even then the total number of Northern Sotho adjectives in the dictionary only amounts to 131 (2.6% of all lemma signs). In contrast to existing dictionaries, where the user must 'construct' the adjective him-/herself, or where a haphazard list of options is given, an explicit approach was followed in the two sides of the dictionary. An example will make this clear. In Kriel et al. (1989⁴) one finds for 'black':

swart, -so (adj.-st.); ... [followed by a selection of combinations]

While Prinsloo and Sathekge (1996) have:

black ntsho, moso, baso, leso

Compare this to the treatment of 'black' in the dictionary under discussion:

black *** adjective, noun

- I adjective **[blacker, blackest]** ⇒cl. 1 **yo moso** Her father was a tall, **black** man. • *Tatagwe e be e le monna yo motelele yo moso.* ⇒cl. 2 **ba baso** This can help **black** women, so that they have something to plant with. • *Se se ka thuša basadi ba baso gore ba be le se sengwe seo ba bjalago ka sona.* ⇒cl. 3 **wo moso** My **black** hair is beautiful. • *Moriri wa ka wo moso o botse.* ⇒cl. 4 **ye meso** She was wearing a dress with **black** and white stripes. • *O be a apere roko ya methaladi ye meso le ye*

mešweu. ⇨cl. 5 **le leso** A fire was made, it was lit and the smoke drifted upwards; it was like a **black** cloud. • *Gwa gotšwa mollo, ya tshungwa, muši wa kuelela godimo e le leru le leso fela.* ⇨cl. 6 **a maso** The sky was covered in **black** clouds that were boiling like the waves of the sea. • *Legodimo le apere maru a maso a a bilogago bjalo ka maphoto a lewatle.* ⇨cl. 7 **se seso** That **black** car is moving. • *Sefatanaga sela se seso se a sepela.* ⇨cl. 8, cl. 10 **tše ntsho** She was wearing **black** shoes (on her feet). • *Kua maotong o be a apere dieta tše ntsho.* ⇨cl. 9 **ye ntsho** Please write in **black** ink. • *Hle ngwalang ka enke ye ntsho.*

II noun [^{pl.} **blacks**] ⇨ **boso**

◇ **black with white back** ⇨cl. 9 **ye kgwadi** ⇨cl. 10 **tše kgwadi** (*said of a bull*)

As can be seen, not only are full forms given (unlike in Kriel et al.), but the translation equivalents immediately contain the demonstrative as well (unlike in Prinsloo and Sathekge). This approach was followed for all 'real' Northern Sotho adjectives.

In order to describe nouns in the absence of enough adjective stems, various grammatical constructions are employed in Northern Sotho, as can be seen from the mini-grammar, viz. possessive concord + noun, demonstrative + relative noun, demonstrative + subject concord + Vgo.

Lastly, in the English to Northern Sotho part of the dictionary, frequent English adjectives that have no frequent counterpart in Northern Sotho, are shown in an abbreviated way, by means of the adjective construction that also underlies the treatment of the article for 'black' just shown, viz., for 'black': [DEM +] CP_{so}; [DEM +] ntsho.

The treatment of adjectives is a clear case where one actually tries to sit on two chairs simultaneously, catering for both the English grammar and the Northern Sotho grammar, and mapping one part of speech onto another one across languages. See for a more in-depth discussion De Schryver (2006).

4.7 Pronouns

❶ **No distinction is made between 'she', 'he' and 'it'** in Northern Sotho. Absolute pronouns (and other words that are used as pronouns) are neutral with regard to gender; therefore the absolute pronoun *yena* can mean either 'she' or 'he'; the possessive pronoun [PC +] *gagwe* can mean either 'her(s)' or 'his', and when used as a possessive pronoun, [PC +] *yona* can mean either 'her(s)', 'his' or 'its', for example: *mahlo a yona* 'her/his/its eyes'.

This seventh point in the mini-grammar is comparable to the first, in that it is only mentioned as a result of the pull of the English language. If one had been dealing with, say, a Northern Sotho–Zulu dictionary, there would not have been a need for this point. The above description is self-contained, rests to add that for the translations of the examples in the Northern Sotho to English side of the dictionary, an attempt was undertaken to distribute the genders according to overall corpus statistics. The occurrence of "she" versus "he" is 40/60%, and likewise for "her(s)" versus "his".

4.8 Demonstratives

③ Three basic positions are distinguished for the **demonstratives** of Northern Sotho. Demonstratives are used to indicate the relative distance between the speaker, the person who the speaker is speaking to (the addressee), and the object or person to which the demonstrative refers. These demonstratives can refer to an object that is (a) relatively close to both speaker and addressee (position I), (b) closer to the addressee where addressee and speaker are relatively far apart (position II), and (c) far away from speaker and addressee, who are quite close to one another (position III). See Table 1 for the various possibilities, including variant positions (Ia, Ib, etc.). Note that these positions are also found for the **demonstrative copulatives**.

Although there are only 47 demonstratives that were frequent enough to be entered into the dictionary (0.9% of all the entries), two thirds of these belong to the top two frequency bands (which means they belong to the top 1 000 words of the Northern Sotho language).

A sound treatment of the demonstratives, as well as of the related demonstrative copulatives, is actually especially difficult on the microstructural level. See De Schryver et al. (2004) for an in-depth, article-length discussion of the lexicographic treatment of the demonstrative copulative in Northern Sotho. As shown there, a sound treatment consists of (a) extra guidance on the microstructural level (by means of context or usage notes), and of (b) multiple levels of cross-referencing, including cross-references between an overview table in the extra matter and the dictionary articles themselves. Both of these have been implemented.

4.9 Locative particles

④ Northern Sotho makes use of **locative particles** to give a detailed description of the place in which an action or process takes place. Five locative particles are distinguished, namely **ka, kua, mo, ga and go**. When the particles *ka*, *kua* and *mo* are used, they are followed by a noun with a locative meaning, which is often a noun with the locative suffix *-ng*. The locative particles can also be combined, thus forming **locative bigrams**, consisting of two locative particles each, and **locative trigrams**, which are combinations of three locative particles. Frequent bigrams are *ka kua* and *ka mo*. Both of these bigrams can combine with *ga* and *go*, leading to the formation of the trigrams *ka kua ga*, *ka kua go*, *ka mo ga* and *ka mo go*. See Figure 2 for the relative distribution of these particles, and Figure 3 for a mnemonic to remember which combinations occur.

On average, every one hundredth word in plain Northern Sotho text and speech is an individual locative particle (cf. De Schryver and Taljard 2006: 141-142). Clearly, one cannot write or speak without them. Even some of the bigrams and trigrams are relatively frequent. As with the demonstratives and the demonstrative copulatives, the main difficulty from a lexicographic point of view is to correctly treat the semantics of the locative particles. As such, the main function of this ninth point in the mini-grammar is to serve as a pointer to the particles themselves — unigrams, bigrams and trigrams — where more information about them may be found.²

4.10 Tone

Ⓢ Northern Sotho is a **tone** language, distinguishing two basic tones, namely high (H) and low (L). Every word has its tone pattern, which might change according to the phonological (or sound) environment in which it appears. The tone pattern for the word *mosegare* 'midday' is LHLH, or *mòségaré*. Tone can be used to distinguish between words which are spelled the same, but have different meanings: *anega* (LLL) means 'tell; narrate', whereas *anega* (HLL) means 'hang (something wet)'. It is not the tradition to indicate tone in Northern Sotho texts, and it is not shown in your dictionary either.

As dictionary compilers, we have frequently been asked why we do not indicate tone in our Northern Sotho reference works. The above paragraph summarises the issue. Tolle lege.

Acknowledgements

Gilles-Maurice de Schryver would like to thank *Ghent University* for its continued support of his field trips to South Africa. Many thanks, too, to the Publisher who was willing to embark on this innovative project.

Endnotes

1. For what it is worth, at the time of writing, a *Google* phrase-search for "corpus-based dictionary grammar" (with the quotes) returns zero hits. In comparison, a search for a word as exotic as "automagically" returns over 2.3 million hits (with a nod to D. Joffe).
2. Doing so, the user will surely be rewarded. At *ka mo ga*, for instance, what must be the most stunning corpus example may be found, viz:
ka mo ga *locative particles* ⇒ **on the side of** Megokgo e tletše mahlo ka lebaka la moya, mongwe o etla ka mo ga hlogo mongwe ka thoko yela gomme e gahlana ka morago ga hlogo. • *Tears filled his eyes because of the wind, one going round this side of the head, the other one on that side, meeting at the back of the head.*

References

- Biber, D. et al. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Corréard, M.-H. (Ed.). 2002. *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. Grenoble: EURALEX.
- De Schryver, G.-M. 2006. Compiling Modern Bilingual Dictionaries for Bantu Languages: Case Studies for Northern Sotho and Zulu. Corino, E. et al. (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6–9 settembre 2006/Proceedings XII Euralex International Congress, Torino, Italia, September 6th–9th, 2006*: 515-525. Alessandria: Edizioni dell'Orso.
- De Schryver, G.-M. 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 E. Taljard.** 2006. Locative Trigrams in Northern Sotho, Preceded by Analyses of Formative Bigrams. *Linguistics, An Interdisciplinary Journal of the Language Sciences* 44(1): 135-193.
- De Schryver, G.-M. et al.** 2004. The Lexicographic Treatment of the Demonstrative Copulative in Sesotho sa Leboa — An Exercise in Multiple Cross-referencing. *Lexikos* 14: 35-66.
- De Schryver, G.-M. et al.** 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.
- Google.* 2007. Google Search Engine [online]. Available <http://www.google.com/>.
- Kriel, T.J., E.B. van Wyk and S.A. Makopo.** 1989^a. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.
- Prinsloo, D.J. and R.H. Gouws.** 1996. Formulating a New Dictionary Convention for the Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 16(3): 100-107.
- Prinsloo, D.J. and B.P. Sathekge.** 1996. *New Sepedi Dictionary, English–Sepedi (Northern Sotho), Sepedi (Northern Sotho)–English*. Pietermaritzburg: Shuter & Shooter.
- 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*. London/Glasgow: Collins ELT.

Table 1: Noun class system and concordial agreement in Northern Sotho

Class (cl.)	CP	Example	SC	OC	PC	DEM pos. I	DEM pos. II	DEM pos. III	DEM pos. Ia	DEM pos. Ib	DEM pos. IIb	DC pos. I	DC pos. II	DC pos. IIIa	PRO Abs	PRO Poss	PRO Poss, communal	PRO Quant
1 1a	mo- Ø-	<i>mosadi</i> 'woman' <i>malome</i> 'uncle'	o / a	mo	wa	yo	yoo	yola	yono	yokhwi	yowe	šo	šoo	šole	yena	gagwe	wešo	yohle
2 2b	ba- bo-	<i>basadi</i> 'women' <i>bomalome</i> 'uncles &c'	ba / ba	ba	ba	ba	bao	bale	bano	bakhwi	bawe	šeba	šebao	šebale	bona	bona	bešo / gabo / gabobona	bohle
3	mo-	<i>monwana</i> 'finger'	o / wa	o	wa	wo	woo	wola	wono	wokhwi	wowe	šo	šoo	šole	wona	wona	wešo	wohle
4	me-	<i>menwana</i> 'fingers'	e / ya	e	ya	ye	yeo	yela	yeno	yekhwi	yewe	še	šeo	šele	yona	yona	yešo	yohle
5	le-	<i>lebone</i> 'light'	le / la	le	la	le	leo	lela	leno	lekhwi	lewe	šele	šeleo	šelele	lona	lona	lešo	lohle
6	ma-	<i>mabone</i> 'lights'	a / a	a	a	a	ao	ale	ano	akhwi	awe	šea	šeao	šeale	ona	ona	ešo	ohle
7	se-	<i>selepe</i> 'axe'	se / sa	se	sa	se	seo	sela	seno	sekhwi	sewe	sese	seseo	sesele	sona	sona	sešo	sohle
8	di-	<i>dilepe</i> 'axes'	di / tša	di	tša	tše	tseo	tšela	tšeno	tšekhwi	tšewe	šedi	šedio	šedile	tšona	tšona	tšešo	tšohle
9	N- Ø-	<i>mpša</i> 'dog' <i>hlogo</i> 'head'	e / ya	e	ya	ye	yeo	yela	yeno	yekhwi	yewe	še	šeo	šele	yona	yona	yešo	yohle
10	diN- di-	<i>dimpša</i> 'dogs' <i>dihlogo</i> 'heads'	di / tša	di	tša	tše	tseo	tšela	tšeno	tšekhwi	tšewe	šedi	šedio	šedile	tšona	tšona	tšešo	tšohle
14	bo-	<i>bodulo</i> 'residence'	bo / bja	bo	bja	bjo	bjoo	bjola	bjono	bjokhwi	bjowe	šebo	šeboo	šebile	bjona	bjona	bješo	bjohle
(6)	ma-	<i>madulo</i> 'residences'	a / a	a	a	a	ao	ale	ano	akhwi	awe	šea	šeao	šeale	ona	ona	ešo	ohle

Class (cl.)	CP	Example	SC	OC	PC	DEM pos. I	DEM pos. II	DEM pos. III	DEM pos. Ia	DEM pos. Ib	DEM pos. IIb	DC pos. I	DC pos. II	DC pos. IIIa	PRO Abs	PRO Poss	PRO Poss, communal	PRO Quant
15	go-	<i>go ruta</i> 'to teach'	go / gwa	go	ga	mo	moo	mola	mono	mokhwi	mowe	šego	šegoo	šegole	gona	gona		gohle / gohlegohle
16	fa-	<i>fase</i> 'below'	go / gwa	go	ga	fa	fao	fale	fano	fakhwi	fawe	šefa	šefao	šefale	gona	gona		gohle / gohlegohle
17	go-	<i>godimo</i> 'above'	go / gwa	go	ga	mo	moo	mola	mono	mokhwi	mowe	šego	šegoo	šegole	gona	gona		gohle / gohlegohle
18	mo-	<i>morago</i> 'behind'	go / gwa	go	ga	mo	moo	mola	mono	mokhwi	mowe	šemo	šemoo	šemole	gona	gona		gohle / gohlegohle
N-	N- Ø-	<i>ntle</i> 'outside' <i>pele</i> 'in front'	go / gwa	go	ga	mo	moo	mola	mono	mokhwi	mowe	šemo	šemoo	šemole	gona	gona		gohle / gohlegohle
(24) <i>ga-</i>	ga-	<i>gare</i> 'middle'	go / gwa	go	ga	mo	moo	mola	mono	mokhwi	mowe	šemo	šemoo	šemole	gona	gona		gohle / gohlegohle

Table 2: Corresponding core information for first and second persons

	SC	OC	PRO Abs	PRO Poss	PRO Poss, communal
1p sg	ke / ka	N-	nna / nnaena	ka	
1p pl	re / ra	re	rena	rena	gešo / gaborena
2p sg	o / wa	go	wena	gago	
2p pl	le / la	le	lena	lena	geno

Notes for Tables 1 and 2:

- All the words printed in bold in Tables 1 and 2 belong to the top-frequency words in Northern Sotho, and are thus included in your dictionary. In order to complete the table, the other forms (in non-bold) have been added, as you may also read or hear them, even though their frequency is low.
- For the N- in Table 2, see the discussion of the verbal prefixes on page X, point ④, on the object concord of the first person singular *n-/m-* 'me'.

Abbreviations used in Table 1, as well as throughout your dictionary:

CP	class prefix	Ø-	zero prefix
DC	demonstrative copulative	1p sg	first person singular
DEM	demonstrative	1p pl	first person plural
OC	object concord	2p sg	second person singular
PC	possessive concord	2p pl	second person plural
PRO	pronoun (absolute, possessive, communal possessive, quantitative)	cl.	class
SC	subject concord	N-	class prefix <i>n-</i> or <i>m-</i>
		pos.	position

Table 3: Single verbal extensions (most frequent ones only, in order of frequency)

Verbal extension	Name	Meaning
-ile	perfect	Indicates that an action was carried out in the past, or that someone/something is in a specific state. English past tenses and past participles are often translated by verbs with the extension <i>-ile</i> in Northern Sotho.
-(i)wa	passive	Adds the meaning of 'be/being' to that of the verb.
-ela	applicative	Adds the meaning of 'to', 'for', 'on behalf of', 'in/to/from the direction of' to that of the verb.
-iša	causative	Adds the meaning of 'cause to', 'help', 'make' to that of the verb.
-ega	neuter-passive	Adds the meaning of 'be/become' to that of the verb.
-ana	reciprocal	Adds the meaning of 'each other / one another' to that of the verb.
-ol, -oll	transitive-reversive	Indicates that the action or state expressed by the basic verb has been reversed. In this way, antonymous verb pairs are formed.

Table 4: Multiple verbal extensions (most frequent ones only, in order of frequency)

Combination of verbal extensions	Example
-ile + -(i)wa	<i>biditšwe</i> 'was/were called' < <i>bitša</i> 'call'
-ela + -(i)wa	<i>agelwa</i> 'be built for' < <i>aga</i> 'build'
-iša + -(i)wa	<i>dirišwa</i> 'be used' < <i>dira</i> 'do'
-ela + -ile	<i>thabetše</i> 'be happy about' < <i>thaba</i> 'be happy'
-ela + -iša	<i>fetetša</i> 'infect; pass on to' < <i>feta</i> 'pass'
-ela + -ela	<i>tsenelela</i> 'penetrate' < <i>tsena</i> 'enter'
-iša + -ela	<i>tiišetša</i> 'strengthen' < <i>tia</i> 'be strong'
-iša + -ile	<i>feditše</i> 'finished' < <i>fela</i> 'finish'
-ega + -ile	<i>diregile</i> 'happened' < <i>dira</i> 'do'

Figure 1: Graphical representation of the frequency of the different verbal extensions (in %)

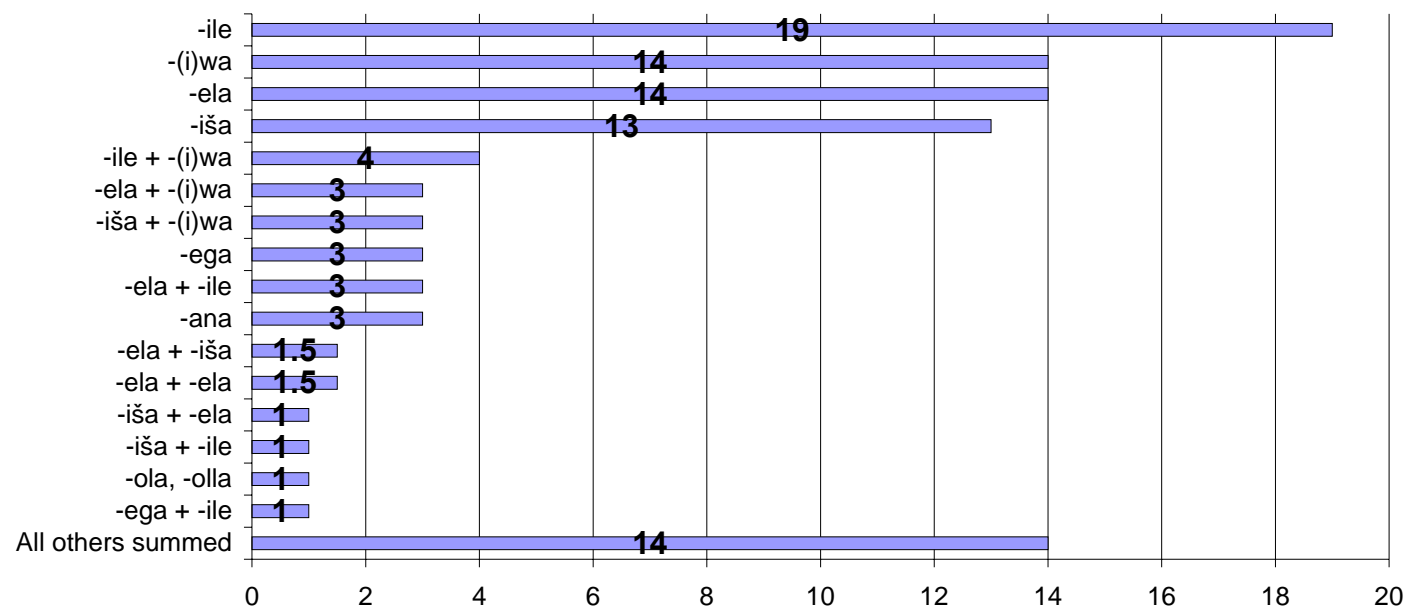


Figure 2: Graphical representation of the frequency of the locative particles

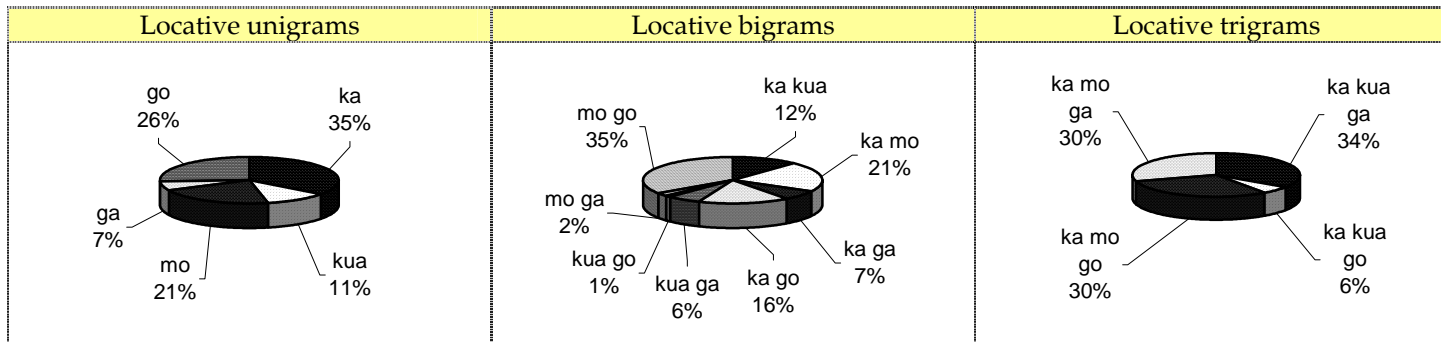
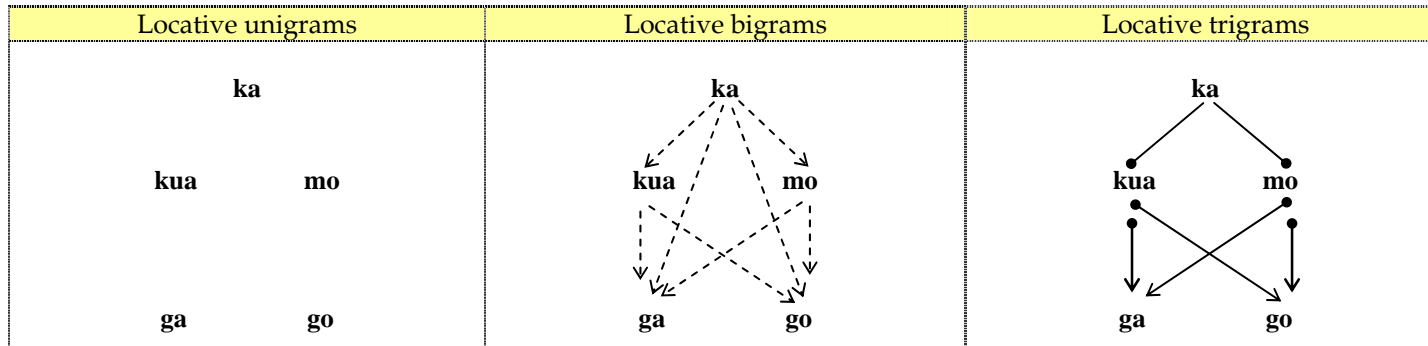


Figure 3: Mnemonic for the possible combinations of locative particles



Terminologiebestuur in Suid-Afrika. 'n Beoordeling van drie werkmodes*

Dina Ferreira, *Skool vir Inligtings- en Kommunikasietegnologie, Sentrale Universiteit vir Tegnologie, Bloemfontein, Republiek van Suid-Afrika*
(dferreir@cut.ac.za)

en

Theodorus du Plessis, *Eenheid vir Taalbestuur, Universiteit van die Vrystaat, Bloemfontein, Republiek van Suid-Afrika*
(dplesslt.HUM@mail.uovs.ac.za)

Opsomming: Die verampteliking van elf tale in Suid-Afrika het tot gevolg dat daar 'n steeds groter wordende behoefte aan die ontwikkeling, die uitbouing en die bestuur van terminologie asook die verspreiding van terminologiese inligting in al hierdie tale bestaan. Hierdie artikel beoordeel drie terminologiese werkmodes wat tans suksesvol gebruik word by terminologiebeplanning, terminologiekoördinerings- en terminologiebestuur.

Deur die navorsing is daar vasgestel dat terminologiebestuur in verskillende omgewings verskillend aangepak word, deur verskillende persone geadminestrer word en verskillende produkte tot gevolg het. Die doelwitte met die versameling en disseminering van terminologie het dus 'n baie belangrike invloed op die werkmodes wat by terminologiebestuur binne 'n spesifieke omgewing gebruik word.

Die drie werkmodes staan bekend as die linguisties georiënteerde werkmodes, die vertalingsgeoriënteerde werkmodes en die taalbeplanningsgeoriënteerde werkmodes. Die implikasies van die verskillende werkmodes word binne die Suid-Afrikaanse situasie ondersoek om te bepaal watter een tans die beste opsie vir terminologieontwikkeling en terminologiebestuur in Suid-Afrika bied.

Sleutelwoorde: AMPTELIKE TALE, BRONTAAL, DOELTAAL, GEMARGINALISEERDE TALE, INHEEMSE TALE, LINGUISTIES GEORIËNTEERDE WERKModes, MEERTALIGE KOMMUNIKASIE, ONTWIKKELDE TALE, ONTWIKKELENDE TALE, TAALBEPLANNING, TAALBEPLANNINGSGEORIËNTEERDE WERKModes, TAALONTWIKKELING, TERMINOLOGIE, TERMINOLOGIEBESTUUR, TERMINOLOGIEONTWIKKELING, TERMINOLOGIEWERK, VERTALINGSGEORIËNTEERDE WERKModes, WÊRELDTAAL

Abstract: Terminology Management in South Africa. A Review of Three Working Methods. The officialising of eleven languages in South Africa has resulted in the

* Hierdie artikel is 'n verwerking en aanvulling van 'n gedeelte uit 'n D.Phil.-proefskrif *Terminologiebestuur in Suid-Afrika met spesifieke verwysing na die posisie van historiese ingekorte tale* wat in 2002 onder die promotorskap van prof. L.T. du Plessis aan die Universiteit van die Vrystaat, Bloemfontein, Republiek van Suid-Afrika, voltooi is..

ever increasing need for the development, the elaboration and the management of terminology as well as the distribution of terminological information in all of these languages. This article reviews three terminological working methods at present successfully used in terminology planning, terminology coordination and terminology management.

It was established through the research that terminology management is dealt with differently in different environments, administered by different people and results in different products. The aims with the collection and dissemination of terminology thus have a very important influence on the working method used for terminology management within a specific environment.

The three working methods are known as the linguistically oriented working method, the translation oriented working method and the language planning oriented working method. The implications of the different working methods within the South African situation are evaluated to establish which one will at present provide the best option for terminology development and terminology management in South Africa.

Keywords: DEVELOPED LANGUAGES, DEVELOPING LANGUAGES, INDIGENOUS LANGUAGES, LANGUAGE DEVELOPMENT, LANGUAGE PLANNING, LANGUAGE PLANNING ORIENTED WORKING METHOD, LINGUISTICALLY ORIENTED WORKING METHOD, MARGINALISED LANGUAGES, MULTILINGUAL COMMUNICATION, OFFICIAL LANGUAGES, SOURCE LANGUAGE, TARGET LANGUAGE, TERMINOLOGY, TERMINOLOGY DEVELOPMENT, TERMINOLOGY MANAGEMENT, TERMINOLOGY WORK, TRANSLATION ORIENTED WORKING METHOD, WORLD LANGUAGE

1. Inleiding

Wanneer daar van taalontwikkeling in die nege amptelike Afrikatale¹ van Suid-Afrika gepraat word, word hoofsaaklik verwys na taalkorpusbeplanning met die oog op nuwe funksieaanwending vir die amptelike taalgebruiksdoeime. Kaplan en Baldauf (2003) sluit hieronder leksikale modernisering, stilistiese modernisering, internasionalisering en renovering in. Laasgenoemde aksie behels onder meer terminologiese unifikasie wat gekoppel kan word aan terminologieontwikkeling en terminologiebestuur. Taalontwikkeling loop saam met taalstatusbeplanning soos die verampteliking van die aangewese tale, 'n aspek van taalbeleidsbeplanning waaraan heelwat aandag in Suid-Afrika geskenk word. Trouens, vir die ontwikkeling van nuwe, onderontwikkelde amptelike tale is gelyklopende taalstatusbeplanning onvermydelik. Die omgekeerde is natuurlik ook waar — sonder taalontwikkeling word dit moeilik om die oogmerke van verampteliking te verwesenlik. 'n Oormatige aksentuering van taalstatusbeplanning kan dus taalontwikkeling kortwiek.

Terminologieontwikkeling en terminologiebestuur vorm derhalwe onontbeerlike onderdele van taalbeplanning binne die nuwe Suid-Afrikaanse taalbestel. Die model vir terminologiebestuur wat tot onlangs toegepas is, is na aanleiding van 'n nasionale vaktaalondersoek in die middel van die tagtigerjare van die vorige eeu ontwikkel. Hierdie model, ontwerp deur Cluver en Scheffer (1984), was egter afgestem op 'n tweetalige taal- en terminologiebestel waar die aksent op die ontwikkeling van Afrikaans geval het en was hoofsaaklik gemik

op vakkundiges in spesifieke vakgebiede om sodoende Afrikaans as taal van die wetenskap en tegnologie te bevorder. Die werkmethode wat in hierdie model gevolg is, staan bekend as die sogenaamde georiënteerde werkmethode. In die lig van die groeiende behoefte aan terminologie in die nege bykomende amptelike tale na 1994, ontstaan die vraag of die betrokke model en sy gepaardgaande werkmethode tans nog relevant is.

Met hierdie artikel word 'n beoordeling van drie terminologiese werkmodes wat op die oomblik suksesvol gebruik word by terminologiebeplanning, -koördinering en -bestuur beoog. Die implikasies van die verskillende werkmodes vir terminologiebestuur binne die Suid-Afrikaanse situasie word ook ondersoek ten einde te kan bepaal watter werkmethode tans die beste opsie vir terminologieontwikkeling en -bestuur in Suid-Afrika bied.

2. Terminologie en terminologiebestuur

2.1 Terminologie

Die polisemie van die term "terminologie" is welbekend. In kontemporêre gebruik word daar gewoonlik tussen drie gebruike van die term onderskei (Sager 1990: 3):

- (a) terminologie as teorie, 'n aanwending wat nodig is om die vooronderstellinge, argumente en gevolgtrekkings tussen konsepte en terme te verantwoord;
- (b) terminologie as aktiwiteit, 'n aanwending wat nodig is om te verwys na die tegnieke en metodes by die versameling, beskrywing en aanbied van terme; en
- (c) terminologie as die woordeskat van 'n spesifieke vakgebied.

Sedert die publikasie van die Internasionale Standaardiseringsorganisasie (ISO) se ISO-standaard 1087, verwys die eerste onderskeiding hierbo na die terminologiewetenskap. ISO 1087 (1990) beskryf die terminologiewetenskap as:

The scientific study of the concepts and terms found in special languages.

Die terminologiewetenskap word erken as 'n aparte wetenskaplike dissipline wat verband hou met die wetenskaplike studie van die konsepte en terme wat in spesiale tale² gevind word.

Die tweede onderskeiding het betrekking op terminologiewerk wat gebaseer word op die beginsels van die terminologiewetenskap. Die betrokke gebruik van die begrip verwys na alle aktiwiteite wat verband hou met die voorbereiding, versameling, prosessering (sistematisering) en aanbieding van terminologiese inligting aan die hand van neergelegde beginsels en metodes. By terminologiewerk is dit die terminoloog se doelwit om ondubbelsinnige terme vir deeglik omskrewe konsepte daar te stel. Om hierdie doelwit te bereik, is dit

nodig om die presiese betekenis van terme te bepaal om sodoende taalgebruikers in staat te stel om sulke terme op 'n universeel aanvaarde wyse te verstaan en te gebruik.

Die derde onderskeiding hierbo het betrekking op die terminologie of terminologieë (dit wil sê vakjargon) van 'n spesifieke taalgebruiksdomein of vakgebied. Volgens hierdie gebruik is terminologie die produk wat gegenereer word deur die terminologiepraktyk. Die term verwys in hierdie geval na die termversameling van 'n spesifieke vakgebied.

By al drie die bogenoemde onderskeidings vorm terminologie derhalwe die essensiële basis vir:

- die ordening van kennis;
- die oordrag van hierdie kennis, vaardighede en tegnologie;
- die oordrag van vakinligting van een taal na 'n ander;
- die formulering van vakinligting;
- die onttrekking en samevatting van vakinligting; en
- die herwinning van gebergde vakinligting.

Vir die doeleindes van hierdie studie word die derde onderskeiding gebruik. Daar word van die veronderstelling uitgegaan dat terminologie noodsaaklik is om professionele kommunikasie in die moedertaal moontlik te maak. Lede van 'n spesifieke spraakgemeenskap kan slegs by wetenskaplike en ander ontwikkeling bybly indien hulle genoegsame, gepaste en gesistematiseerde terminologie in hulle moedertaal tot hulle beskikking het. Terminologiewerk in die moedertaal vorm dus deel van die dinamiek van vitale spraakgemeenskappe.

2.2 Terminologiebestuur

Galinski en Budin (1993: 209) beskryf terminologiebestuur as 'n konsep parallel aan inligtingsbestuur.

The term 'terminology management' can be considered a special kind of 'information management' that focusses on structuring, storing, exchanging, disseminating and using terminological information for text function (including dictionaries etc.). It is thus a wider and more complex activity than terminography viz. the recording of terminological information.

Beide inligtingsbestuur en terminologiebestuur is onmisbare instrumente binne 'n omgewing waar daar gewerk word met kwaliteitbestuur. Met terminologiebestuur word onder meer beoog om tred te hou met die sogenaamde inligtingsontploffing en om die toename in die hoeveelheid inligting en gepaardgaande terminologie te bestuur, te organiseer en te beheer.

Wright en Budin (1997: 2) sien terminologiebestuur as enige doelbewuste manipulerings van terminologiese inligting. Drie aspekte het betrekking, te wete

praktiese terminologiebestuur, die sistematiese versameling van terminologiese inligting en die *ad hoc*-onttrekking van sodanige inligting. Praktiese terminologiebestuur sorg dat daar deur ingenieurs, bioloë en wetenskaplikes voldoen word aan die konstante behoefte aan presiese en ondubbelsinnige taal in die beoefening van hulle professionele beroepe. Nuwe konsepte word deurlopend benoem deur nuwe terme te skep om sodoende kommunikasie te bly moontlik maak. Die sistematiese versameling van terminologiese inligting en die aanbieding daarvan in die vorm van woordeboeke, terminologiese versamelings en terminologiese databasisse is 'n tweede aspek en word gewoonlik behartig deur taalwerkers, taalpraktisyne en persone wat spesifiek vir hierdie taak aangewend word. Die *ad hoc*-onttrekking van terminologiese inligting vir die doeleindes van vertaling en tolking, vir teksprodusering as 'n funksie van tegniese kommunikasie, en vir inligtingsbestuur is die derde aspek waarna hierdie outeurs verwys. Hierdie aspek het betrekking op die terminologieverbruiker.

Cabré (1999: 20) omskryf terminologiebestuur as die organisering van terminologiewerk en wys daarop dat 'n omvattende beskrywing van die organisering van terminologiewerk begin by 'n analise van al die fasette waaruit dit bestaan. Daar moet volgens haar ook vasgestel word hoe elk van hierdie fasette georganiseer word en wie vir elkeen verantwoordelik is. Sy benadruk die feit dat daar binne 'n spesifieke sosio-politieke omgewing 'n gespesialiseerde terminologiesentrum behoort te bestaan wat al hierdie organisatoriese fasette hanteer.

Uit bogenoemde is dit duidelik dat terminologiebestuur noodsaaklik is om terminologiese inligting te organiseer en te koördineer as voorwaarde vir die effektiewe verspreiding van kennis en die daarstel van hoë gehalte terminologiese inligting. Die primêre doelwit met terminologiebestuur behels derhalwe die skep van werks- en ander omstandighede om 'n hoë gehalte terminologie-uitset te kan lewer. Terminologiebestuur plaas met ander woorde 'n hoë premie op produktiwiteit en gehalte.

3. Werkmetodes by terminologiebestuur

Terminologiebestuur word in verskillende omgewings verskillend aangepak en deur verskillende persone geadmistreer. Die doelwitte met die versameling en disseminering van terminologie het 'n baie belangrike invloed op die werkmetode wat by terminologiebestuur binne 'n spesifieke omgewing gebruik word.

Volgens Auger (1988) en Cabré (1999) bestaan daar drie werkmetodes waarvolgens terminologiewerk tans suksesvol beplan, gekoördineer en bestuur word:

- die sogenaamde linguïsties georiënteerde werkmetode, waar die fokus val op konsepte en die standaardisering van terme en begrippe — terminologie word aangepas om 'n linguïstiese sisteem te dien;

- die sogenaamde vertalingsgeoriënteerde werkmodes, waar die fokus val op databasiskepping — terminologie word ontwikkel om vertaling te dien; en
- die sogenaamde taalbeplanningsgeoriënteerde werkmodes, waar die fokus val op sistematiese tussenbeidetreiding ten einde die status van 'n (onstabiele) taal te verander — terminologie word ontwikkel om 'n taalbeplanningsagenda te dien.

Hierdie drie werkmodes sal saaklik toegelig word.

3.1 Die linguisties georiënteerde werkmodes

By die linguisties georiënteerde werkmodes word terminologie benader as 'n reeks ooreengekome terme vir 'n spesifieke spraakgemeenskap. Terminologie word geskep en bestuur om misverstande binne hierdie spraakgemeenskap te voorkom en om kommunikasie te vergemaklik. Hierdie werkmodes gaan van die standpunt uit dat terminologiewerk by konsepte begin en dat daar vanaf 'n definisie gewerk word na 'n geskikte term toe. Die linguisties georiënteerde werkmodes het hoofsaaklik ontstaan uit 'n behoefte onder tegnisi en wetenskaplikes om die terminologie binne hulle spesialiteitsvelde te standaardiseer om sodoende effektiewe kommunikasie en kennisverspreiding onder spesialiste te verseker.

Drie klassieke skole van terminologie onderskryf die linguisties georiënteerde werkmodes, te wete die Weense Skool van Terminologie, die Tsjeggiese Skool van Terminologie en die Sowjetse Skool van Terminologie.

Die Weense Skool van Terminologie is moontlik die bekendste en die invloedrykste. Die benadering van hierdie skool word gebaseer op die werk van Eugene Wüster getiteld *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie* (Wüster 1979).

Baie lande in sentraal en Noord-Europa (Oostenryk, Duitsland, Noorweë, Swede en Denemarke) werk binne die raamwerk van die Weense Skool van Terminologie, waar vakspesialiste self verantwoordelik is vir die ontwikkeling van gespesialiseerde terminologieë.

Die Tsjeggiese Skool het ontstaan as resultaat van die funksionele linguistiese benadering wat deur die Praagse Skool van Linguistiek ontwikkel is. Die hoofdoelwit met die funksionele linguistiek is dat kommunikatiewe doelwitte binne die kommunikasie-opset funksioneel moet wees.

Die Sowjetse denkskool het baie raakpunte met die Weense Skool. Die Sowjetse Skool is hoofsaaklik geïnteresseerd in die standaardisering van konsepte en terme in die lig van probleme wat verband hou met meertaligheid in die vorige Sowjetunie (nou die Statebond van Onafhanklike State).

Die linguisties georiënteerde werkmodes by terminologiebestuur konsentreer op konsepte en lei terminologiewerk in die rigting van term- en konseptestandaardisering. Vakkundiges en spesialiste wat hierdie werkmodes volg, is self verantwoordelik vir gespesialiseerde terminologieë.

Standaardisering vorm 'n primêre komponent by die linguisties georiënteerde werkmethode waar spesiale kommunikasie sonder dubbelsinnigheid veronderstel dat elke benaming gekoppel word aan 'n enkele konsep en elke konsep deur 'n enkele term.

Twee internasionale liggame, die Internasionale Elektrotegniese Kommissie (IEC) en die Internasionale Organisasie vir Standaardisering (ISO), het as primêre doelwit die ontwerp van internasionale riglyne vir standaardisering van terminologiese inligting. Hulle volg hierdie werkmethode.

Die tegniese Komitee 37, Terminology and Other Language and Content Resources van die Internasionale Organisasie vir Standaardisering (ISO/TC 37), voorsien standaarde vir terminologiedata vir internasionale standaardiseerders. ISO/TC 37 het in 1992 'n dokument gepubliseer wat bekend staan as die *International Standards — Preparation and Layout* (ISO 10241: 1992). Hierdie dokument stel riglyne vas vir gebruik by die voorbereiding en uitleg van internasionale standaarde.

As internasionale organisasie maak ISO staat op die samewerking van nasionale standaardiseringsliggame. Die meeste van die wêreld se onderskeie nasionale standaarde oor terminologiebeginsels stem ooreen met die ISO-standaarde, alhoewel nasionale standaarde soms meer gedetailleerd is omdat kenmerkende eienskappe van spesifieke tale in ag geneem kan word. Om standaardiseringsprojekte van stapel te stuur, word Tegniese Komitees geskep waarin die nasionale standaardiseringsliggame verteenwoordig word. Tans het ISO ongeveer 170 nasionale komitees en onderskryf nagenoeg 90 lande ISO se standaarde.

Die algemene motivering vir standaarde is om uniforme metodes in terminologiebestuur te verskaf:

- om standaardiseringsliggame by te staan om hulle terminologiese bestuur;
- om die kommunikasie van die inhoud en wêreldwye gebruik van tegniese tale te bevorder;
- om vakkundiges by te staan om meer effektiewe terminologiebestuur te verkry; en
- om makliker toegang tot verskillende velde van kennis in internasionale standaarde te verskaf.

3.2 Die vertalingsgeoriënteerde werkmethode

By die vertalingsgeoriënteerde werkmethode word terminologiebestuur gebruik om:

- terminologie-ekwivalente in verskillende tale te versamel en te sistematiseer;

- terminologie-ekwivalente in verskillende tale geredelik oor tyd heen te gebruik; en
- meer effektief met terminologie-ekwivalente in verskeie tale te werk.

Die vertalingsgeoriënteerde werkmodes word gevolg binne die konteks van institusioneel tweetalige of multitalige provinsies of lande, byvoorbeeld in Quebec asook Wallonië, die Franssprekende gedeelte van België. Die vertalingsgeoriënteerde werkmodes het ontwikkel in die lande waar koloniale tale gebruik word. Die terminologieontwikkeling en terminologiebestuur wat deur hierdie lande onderneem word, is hoofsaaklik eensydig aangesien die kriteria wat opgestel word vir die bevordering en uitbreiding van veral Frans (en ook Engels) nie toegepas word op die inheemse tale van vorige kolonies nie.

Die vertalingsgeoriënteerde werkmodes behels dat vertaalekwivalente in al die betrokke tale beskikbaar gemaak word om vertalers in staat te stel om kommunikasie te vergemaklik (vgl. Cabré 1999: 15). Dit is 'n metode wat die basis vorm van terminologiese aktiwiteite wat deur veeltalige lande en veeltalige internasionale organisasies onderneem word. Vertalers identifiseer terminologiese inligting binne tekste na aanleiding van vooraf vasgestelde kriteria. Die inligting word dan ondersoek en gedokumenteer deur van betroubare en gesaghebbende bronne gebruik te maak. Vertalers werk saam met vakkundiges om maniere te vind waarop terminologiese inligting doelmatig geëkspeereer, vertaal, opgeteken en gedissemineer kan word.

Voorbeelde van instansies wat die vertalingsgeoriënteerde werkmodes volg, sluit in die Verenigde Nasies, Unesco self en die Europese Unie. Die federale Kanadese regering val ook onder hierdie groep en is aktief betrokke by terminologie-ontwikkeling. Omdat Kanada 'n tweetalige land is, is die Kanadese regering verantwoordelik vir die bevordering en ontwikkeling van die taalregte van al sy burgers.

By die vertalingsgeoriënteerde werkmodes is die doelwit met terminologiebestuur om te verseker dat terminologiese ekwivalente deur vertalers in die verskillende tale gebruik kan word. So word 'n bydrae gelewer tot die gehalte van vertaalde tekste deur 'n tipe universaliteit te verseker en kommunikasie te vergemaklik.

ISO TC 37/SC2/WG2 het 'n dokument *Translation-oriented Terminography* (ISO 12616: 2002) gepubliseer wat riglyne ontwerp vir die uitvoering van die vertalingsgeoriënteerde werkmodes by terminologiebestuur. ISO 12616 (2002) verskaf normatiewe riglyne wat vertalers in staat stel om terminologiese inligting te versamel, te sistematiseer en vinnig en maklik te onttrek. Bykomend tot inligting oor die spesiale kenmerke van vertalingsgeoriënteerde terminografie, stel die standaard ook prosedures vas vir die bestuur van terminologiese inligting binne 'n gerekenariseerde omgewing. Verder identifiseer die standaard ook noodsaaklike en opsionele datakategorieë wat benodig word binne die konteks van die vertalingsproses.

3.3 Die taalbeplanningsgeoriënteerde werkmethode

Die taalbeplanningsgeoriënteerde werkmethode by terminologiebestuur het in ongeveer 1960 posgevat en is aanvanklik ontwikkel om die gebruik van minderheidstale te ondersteun. Die onderliggende gedagte is om die gebruik van 'n onstabiele taal te verander deur middel van sistematiese en strategiese tussenbeidetrede, onder meer deur middel van die regte wetgewing en ander doelgerigte maatstawwe. Intensiewe pogings word aangewend om tale wat gewoonlik terminologies onderontwikkel of slegs gedeeltelik ontwikkel is, as instrumente vir professionele kommunikasie uit te bou. Voorbeelde van tale wat deur die taalbeplanningsgeoriënteerde werkmethode beïnvloed kan word of reeds beïnvloed is, is Afrikaans (tydens die tweetalige bestel), die Afrikatale, Hebreeus, Arabies en Hindi. In die spraakgemeenskappe waar hierdie tale gepraat word, word die gebruikers steeds grootliks verplig om binne die tegniese en wetenskaplike registers van Engels of Frans te opereer aangesien hulle tale nie oor die nodige terminologie beskik om op alle vakgebiede en -domeine aangewend te word nie. Om 'n verandering teweeg te bring, moet die betrokke tale voldoende en koherente terminologie hê om professionele kommunikasie op alle gebiede moontlik te maak (vgl. Cabré 1999: 14). Taalwetgewing is een manier om dit te bewerkstellig.

Hierdie werkmethode word primêr geïnspireer deur die taalbeplanningsbeleid binne 'n spraakgemeenskap. 'n Goeie voorbeeld van 'n regering wat van die taalbeplanningsgeoriënteerde werkmethode gebruik maak is dié van Quebec. 'n Beleid is in 1970 ingestel om die posisie van Frans in Kanada te bevorder deur die uitbreiding daarvan in alle gebruiksfere te verseker (Cabré 1999: 14).

Regerings raak om verskillende redes betrokke by sodanige taalbeplanning, onder meer om kulturele onafhanklikheid in postkoloniale lande te bewerkstellig asook om nediskriminerende politiek in veeltalige lande of organisasies te beklemtoon.

Die taalbeplanningsgeoriënteerde werkmethode blyk die relevantste te wees in gemeenskappe waar daar 'n taalbeleid bestaan om die status van inheemse tale te verander. In hierdie gemeenskappe word twee of meer tale gewoonlik komplementêr tot mekaar aangewend op verskillende gebiede van kommunikasie. Gewoonlik is een van die tale 'n wêreldtaal, byvoorbeeld Engels. Hierdie taal word dan as die brontaal gebruik waarin terme en definisies versamel word. Vertaalekwivalente moet gevolglik in die inheemse taal verskaf word.

Hierdie benadering tot terminologiebestuur het 'n tweeledige oogmerk, een wat gemik is op die standaardisering van ekwivalente terminologie (terminologiestandaardisering) en een wat gemik is op die standaardisering van 'n agtergestelde taal (taalstandaardisering). Taalstandaardisering verwys uiteraard nie na terminologiestandaardisering waarby die ISO-komitee betrokke raak en waar terminologie gelys, omskryf en hoofsaaklik net deur vakspesialiste gebruik word nie. Terminologiebestuur volgens die taalbeplanningsgeoriënteerde werkmethode het die ontwikkeling en bevordering van 'n onderont-

wikkelde taal ten doel sodat die taal vir hoë funksies in die samelewing aangewend kan word. Nuwe terme mag geskep word waar daar nie bekende terme bestaan nie, maar selfs belangriker is dat terme versamel, gesistematiseer en voorberei moet word vir gebruik in verskillende vakgebiede en domeine. Die taalbeplanningsgeoriënteerde werkmodes verseker dat die taalgebruikers in staat gestel word om hulle tale in verskillende funksies aan te wend. Ferguson (1972: 72) se klassieke onderskeiding tussen taalgebruik vir die hoë funksies en taalgebruik vir die lae funksies binne 'n samelewing het hier betrekking. Hoë-funksietaalgebruik word geassosieer met die gebruik van 'n taal in domeine soos wetgewing, regerings- en publieke administrasie, onderwys en justisie, sowel as in die nywerheid en handel, kultuur, die media, ensovoorts. Deur 'n taal in hoë funksies te gebruik, word die status van daardie taal verhoog, wat die verhouding tussen taalstatus en taalgebruik beklemtoon.

By die taalbeplanningsgeoriënteerde werkmodes word terminologie wat ingevoer word uit tale wat in terminologiedominante lande gebruik word, vervang met terminologie in die inheemse tale. Dit word gerealiseer deur 'n taalbeleid en taalbeplanningsprogram wat die uitbreiding en bevordering van die inheemse tale en noodwendig ook hulle ontwikkeling ten doel het.

3.4 'n Vergelyking van die drie werkmodes by terminologie

Uit die oorsig tot hiertoe kan afgelei word dat die oogmerke by elke werkmodes verskil en dat die resultate en produkte wat gelewer word ook kan verskil. Die belangrikste oogmerke, vereistes, rolspelers en produkte van die drie werkmodes word opsommenderwys in Tabel 1 aangebied.

Die verskille tussen die drie werkmodes ten opsigte van die beplanning, koördinering en bestuur van terminologiebronne kan waargeneem word by die verskillende oogmerke met die terminologie-inligting, die vereistes wat noodsaaklik is, die rolspelers wat betrokke is en die produkte wat deur die verskillende werkmodes beskikbaar gestel word.

Die oogmerke met die linguisties georiënteerde werkmodes behels term- en konsepstandaardisering (gewoonlik eentalig) om sodoende effektiewe en presiese kommunikasie tussen vakkundiges en spesialiste te bewerkstellig. Die vereistes vir hierdie werkmodes is die aktiewe betrokkenheid van moedertaalvakkundiges by die ontwikkeling van wetenskaplike en tegniese taal. In die geval van terminologiewerk wat gemik word op die standaardisering van konsepte en terme, word die proses om 'n standaard te kies of om 'n konsep te benoem gewoonlik behartig deur tegniese komitees van vakspecialiste. Die taak van hierdie tegniese komitees behels die kies van een van die beskikbare terme in die geval van alternatiewe om 'n enkele nasionale of internasionale benaming vir elke konsep vas te stel. Standaardisering by hierdie werkmodes vind gewoonlik plaas deur van ISO-aanbevelings gebruik te maak met presiese definisies en min oorbodige inligting. Die ordening is sistematies, aangesien die gebruikers spesialiste en vakkundiges is wat die vakgebied ken.

WERKMETODE	OOGMERKE	VEREISTES	ROLSPELERS	PRODUKTE
Linguisties georiënteerde werkmethode	Term- en konsepstandaardisering wat effektiewe en presiese kommunikasie tussen vakkundiges en spesialiste verseker	Aktiewe moedertaalvakkundiges	Standaardiseringsliggame Eentalige tegniese komitees bestaande uit linguïste, vakkundiges en terminoloë	ISO-gestandaardiseerde terminologie-inligting wat kommunikasie tussen vakkundiges fasiliteer
Vertalingsgeoriënteerde werkmethode	Terminologiese ekwivalensie tussen ontwikkelde tale om universaliteit te bevorder en kommunikasie te vergemaklik	Aktiewe moedertaalvakkundiges Beskikbare terminologie in die brontaal Vertaalde korpora Taalwetgewing: gelykstelling van betrokke tale	Meertalige tegniese komitees bestaande uit linguïste, vakkundiges, terminoloë en vertalers	Terminologie-inligting in ontwikkelde tale wat vertaling bevorder en vergemaklik
Taalbeplanningsgeoriënteerde werkmethode	Verhoging van die taalstatus van onderontwikkelde tale deur gepaste taalwetgewing of taalbeleid en ander doelgerigte taalbeplanningspogings (m.a.w. verampteliking), met die klem op taalontwikkeling	Dominante brontaal Brontaalterminologie Taalwetgewing of taalbeleid Taalontwikkelingsplan Skep, versameling, sistematisering en voorbereiding van terme om taalgebruikers in staat te stel om tale vir verskillende funksies te gebruik Vertaling van tekste om dokumentasie vir gebruikers toeganklik te maak	Meertalige linguïste, terminoloë en gebruikers	Terminologie-inligting in onderontwikkelde tale wat die tale vir hoëfunksiegebruik geskik maak

Tabel 1: 'n Opsomming van die werkmodes by terminologiebestuur

In teenstelling hiermee behels die vertalingsgeoriënteerde en die taalbeplanningsgeoriënteerde werkmodes nie net terminologieswerk in slegs 'n enkele taal se wetenskaplike en tegniese domeine nie. Die doelwit is meertalig en ook prakties van aard, naamlik die fasilitering van vertaling of die voorbereiding van terminologiese vir tale waarin daar 'n terminologiestekort bestaan. Hierdie twee denkwyses konsentreer op alle domeine van kennis en menslike aktiwiteite. Tydens die beplanning en koördinerende van terminologieswerk vir hierdie twee denkwyses, moet die belangrikheid van die verskillende kommunikasiekontekste in gedagte gehou word om die gebruikers van die terminologiesinligting van 'n doeltreffende spesiale taal te voorsien.

Die oogmerke met die vertalingsgeoriënteerde werkmodes behels dat terminologiese ekwivalensie tussen ontwikkelde tale moet bestaan om universaliteit te bevorder en kommunikasie te vergemaklik. Die vereistes by hierdie werkmodes is dat taalwetgewing ingevoer word wat die gelykstelling van die betrokke tale bepaal, dat moederdoeltaalvakkundiges aktief betrokke moet wees om beskikbare terminologies in die brontaal doelmatig te ekserpeer en dat vertaalde korpora bestaan waaruit ekwivalente onttrek kan word. By die vertalingsgeoriënteerde werkmodes speel vakspecialiste in die verskillende tale en vertalers in die doeltaal(-tale) 'n baie belangrike rol, aangesien dokumente in die verskillende tale naas mekaar beskikbaar gemaak moet word om sodoende kommunikasie te vergemaklik en misverstande uit te skakel. Vakspecialiste word benodig om genoegsame terminologies in die tale te voorsien en vertalers word in die doeltaal(-tale) benodig om dokumente deur vertaling vir verskillende taalgroepe toeganklik te maak.

In die geval van terminologieswerk wat gedoen word met die oog op taalbeplanning binne 'n gemeenskap, word 'n verskillende doelwit nagestreef. Die oogmerke met die taalbeplanningsgeoriënteerde werkmodes behels 'n verhoging van die taalstatus van onderontwikkelde tale deur gepaste taalwetgewing of taalbeleid en ander doelgerigte taalbeplanningspogings (soos verampteliking). Die klem val hier op taalontwikkeling. Die vereistes by hierdie werkmodes behels dat terminologies in die brontaal moet bestaan en dat taalwetgewing of taalbeleid ingestel moet word om die doeltaal(-tale) te bevorder en te ontwikkel. Wat by hierdie werkmodes van kardinale belang is, is die versameling, sistematisering, skep en voorbereiding van terminologiese in die onderontwikkelde tale om die taalgebruikers van hierdie tale in staat te stel om die tale vir verskillende funksies te gebruik asook die vertaling van tekste vanuit die brontaal te fasiliteer om dokumentasie sodoende toeganklik te maak vir die gebruikers van die onderontwikkelde tale. By hierdie werkmodes speel linguïste en die gebruikers van terminologiesinligting 'n baie belangrike rol: linguïste, aangesien die fokus van die werk op die meer algemene domeine val wat nie noodwendig 'n gespesialiseerde formele struktuur insluit nie en die gebruikers, aangesien hulle verantwoordelik is vir die aanvaarding en verspreiding van die terminologies.

4. Implikasies van die werkmodes vir terminologiebestuur in Suid-Afrika

Die toepasbaarheid van elk van die werkmodes by terminologiebestuur in Suid-Afrika sal vervolgens ondersoek word deur van die kriteria in Tabel 1 gebruik te maak. Hierdeur sal dan vasgestel word in hoeverre die spesifikasies vir die oogmerke, die vereistes en die rolspelers by elke werkmode as veranderlikes aanwesig is by terminologiebestuur in Suid-Afrika.

4.1 Die linguisties georiënteerde werkmode

Vir die linguisties georiënteerde werkmode by terminologiebestuur om effektief te kan wees binne 'n gemeenskap, word reeds ontwikkelde tale benodig waar wetenskaplikes en tegnisiërs as vakkundiges en spesialiste binne 'n spesifieke vakgebied bestaan om korpora in die betrokke taal beskikbaar te stel, waarvandaan die terme en definisies dan geëkskipeer, opgeteken, (volgens ISO-standaarde) gestandaardiseer en gedissemineer kan word.

Die Suid-Afrikaanse taalbestel is veeltalig en slegs een van die amptelike tale, naamlik Engels, kan as 'n reeds ontwikkelde taal beskryf word. Afrikaans is in 'n sekere mate ontwikkel as gevolg van die vorige taalbestel waar dit gelykgestel was aan Engels. Die ander nege amptelike tale daarenteen kan as ontwikkelende tale beskou word.

Alhoewel daar in Afrikaans en by sommige van die Afrikatale velde of domeine mag bestaan waar die linguistiese werkmode met sukses gebruik kan word, is daar oor die algemeen min vakkundiges en wetenskaplikes wat Afrikaans en die Afrikatale aktief in wetenskaplike artikels in geakkrediteerde tydskrifte aanwend. By Afrikaans gaan dit nie noodwenig oor die vaktale wat te kort skiet aan terme wat die publikasie van vakartikels in Afrikaans kortwiek nie, maar eerder die feit dat vakkundiges internasionale blootstelling soek en daarom in Engels publiseer. Engels word hoofsaaklik in hierdie domeine en velde gebruik wat die taal dan tegnisiërs 'n terminologiebrontaal maak. Mouton (2005: 371) het bevind dat die totale aantal geakkrediteerde artikels in die databasis (South African Knowledgebase) tussen 1990 en 2002 90 713 was waarvan 81 565 (89,9%) in Engels, 7 453 (8,2%) in Afrikaans, en 1 791 (2,2%) in ander tale (hoofsaaklik Nederlands, Duits en Frans). Teksproduseerders verkies skynbaar om eerder van Engels wat as 'n wetenskaplike en terminologiese wêreldtaal beskou word, gebruik te maak, aangesien die inligting wat hulle wil beskikbaar stel, volgens hulle redenasie op dié manier 'n wyer teikengroep bereik.

Die feit dat terminologiedata nie ondersteun word deur geskrewe wetenskaplike dokumentasie en tekste in veral die Afrikatale nie het tot gevolg dat die terminologie nie deurlopend in die gemeenskap sirkuleer nie. Derhalwe word terminologie tans nie op 'n wyse aangebied wat vir die gebruiker 'n aanduiding gee van hoe die terminologie in konteks aangewend word en hoe ander gebruikers oor die terminologie voel nie.

Terminologiestandaardisering volgens ISO-standaarde wat deur reeds ontwikkelde tale in hoofsaaklik eentalige lande gebruik word, is met ander woorde nog nie moontlik in al die amptelike tale van Suid-Afrika nie. Afrikaans word hier uitgesluit aangesien ISO-standaarde reeds sedert die vyftigerjare by terminologiestandaardisering gebruik word. Desnieteenstaande blyk die linguïsties georiënteerde werkmodes in hierdie stadium nie 'n geskikte werkmodes te wees waarvolgens terminologie in al die amptelike tale van Suid-Afrika bestuur kan word nie.

4.2 Die vertalingsgeoriënteerde werkmodes

Vir die vertalingsgeoriënteerde werkmodes by terminologiebestuur om effektiwiteit te kan wees, word vereis dat die betrokke tale reeds ontwikkelde tale binne 'n veeltalige taalbestel moet wees en dat terminologie-ekwivalente in al die betrokke tale beskikbaar moet wees om die vertaling van tegniese en wetenskaplike dokumentasie te vergemaklik. Die beskikbaarheid van korpora in beide die bron- en die doeltale is derhalwe by die vertalingsgeoriënteerde werkmodes 'n voorvereiste om die ekserpering van ekwivalente terminologie moontlik te maak en om tweetalige of meertalige terminologiese beskikbaar te stel.

In die vorige taalbestel het die vertalingsgeoriënteerde werkmodes by terminologiebestuur wel goed gewerk, aangesien Engels en Afrikaans feitlik op 'n gelyke vlak van uitbreiding was en hierdie twee tale gelykmatig as tale vir tegniese en wetenskaplike kommunikasie gebruik is. Die keuse van vakgebiede en domeine by terminologiese werk in die vorige taalbestel is bepaal deur nasionale prioriteite — die uitbreiding van Afrikaans ingevolge die tweetaligheidsbeleid.

Die Afrikatale is egter nie in die vorige taalbestel tot op dieselfde vlak van terminologiese ontwikkeling as Afrikaans gebring nie en kan tans as onderontwikkelde tale beskou word wat in terme van die nuwe taalbestel uitgebou moet word.

Die vertalingsgeoriënteerde werkmodes wat gedurende die vorige taalbestel gebruik is om Afrikaans te ontwikkel tot op 'n gelyke vlak met Engels, kon dus nie in die nuwe taalbestel oorgeneem word nie, aangesien Suid-Afrika 'n veeltaligheidsbestel betree het waar die verskillende amptelike tale nie almal dieselfde vlak van ontwikkeling besit nie. Die teikengroep waarvoor terminologiese uitbreiding in die nuwe taalbestel gedoen word, verskil ook in 'n groot mate van die teikengroep van die vorige taalbestel. Die teikengroep is meertalig en die terminologiese doelwit is meer prakties van aard, naamlik die fasilitering van terminologiese vir die tale waarin daar 'n terminologiese tekort bestaan. 'n Meer toepaslike werkmodes moet derhalwe oorweeg word om terminologiese bestuur by die behoeftes en uitdagings van die nuwe terminologiese omgewing aan te pas.

4.3 Die taalbeplanningsgeoriënteerde werkmethode

Vir die taalbeplanningsgeoriënteerde werkmethode by terminologiebestuur om effektief te wees, word die sistematiese en strategiese tussenbeidetreeding deur middel van taalwetgewing of taalbeleid vereis om die taalstatus van minderheidstale te verbeter en hierdie tale dienooreenkomstig te ontwikkel. 'n Taalbeplanningsprogram is dus onontbeerlik.

Die feit dat die Grondwet amptelike erkenning op nasionale vlak verleen aan elf amptelike tale vir Suid-Afrika (Artikel 6(1)), impliseer dat hierdie tale 'n belangrike rol moet speel in die land se sosiale, ekonomiese, kulturele en politieke domeine, met ander woorde dat hierdie tale deur hulle sprekers in hierdie domeine gebruik moet word en dat geen hiërargisering ten opsigte van taal toegepas mag word nie (Du Plessis en Pretorius 2000). Dit bring mee dat die Afrikatale wat vroeër beperkte amptelike funksies verrig het, ontwikkel moet word om diens te doen as amptelike tale (Artikel 6(2), (3), (4)).

'n Beleidsriglyn van die taalklousule van die Grondwet is dat al die amptelike tale gelykheid van aansien moet geniet. Dit beteken dat sekere doelwitte bestaan in terme waarvan al die amptelike tale in enige register en op alle domeine van taalgebruik aangewend kan word. Die Grondwet ken status en funksies aan die Afrikatale toe (legitimisering van die tale), maar die status moet as 't ware verwerf word deur die gebruik van hierdie tale (institusionele van die tale).

Gegewe die vorige taalbestel, is dit nie altyd moontlik om die Afrikatale ten opsigte van al die domeine in hierdie hoë funksies te gebruik nie en in elk geval nie op 'n basis wat gelykheid van aansien verteenwoordig nie, omdat hulle nie genoegsaam ten opsigte van vorm ontwikkel is nie. Intensiewe taalontwikkeling in die Afrikatale is derhalwe nodig, sodat hulle funksies uitgebrei kan word en hulle op alle vlakke van die samelewing as effektiewe kommunikasiemiddels gebruik kan word. Engels word steeds in baie vakgebiede en domeine as kommunikasiemiddel aangewend, aangesien die ander tale nie oor voldoende en koherente terminologie beskik nie.

Engels as wêreldtaal kan derhalwe volgens die taalbeplanningsgeoriënteerde werkmethode gebruik word vir die onttrekking van terme en die saamstel van definisies in die agtergestelde tale. Afrikaans en die ander nege amptelike tale (wat tans prioriteit in terme van ontwikkeling geniet) kan as doeltale funksioneer waarin ekwivalente verskaf moet word.

Van die belangrikste taalbeleidsdokumente wat die beleid en praktyk rondom die taalbeplanningsgeoriënteerde werkmethode ondersteun, word saaklik vermeld:

Die sogenaamde LANGTAG-verslag (DACST 1996) bevat 'n "Opsomming van die belangrikste aanbevelings van LANGTAG" (vgl. DACST 1996: 201-204) wat op die uitbreiding van terminologie en die bestuur van terminologie in Suid-Afrika betrekking het.

-
- As korttermynmaatreëls word voorsien:
 - Die bevordering van die gebruik van ander tale as Engels en Afrikaans in nuwe domeine en in hoërstatusfunksies, bv. by universiteite en teknikons.
 - Die skep van nuwe databasisse en die ouditering van bestaande databasisse via die Pan-Suid-Afrikaanse Taalraad (PanSAT), die Raad vir Geesteswetenskaplike Navorsing (RGN), die Departement Kuns, Kultuur, Wetenskap en Tegnologie (DKKWT), Gemeenskapsgebaseerde Organisasies (GBO's), Nieregeringsorganisasies (NRO's) en ander organe oor relevante taalverwante aangeleentehde sodat navorsing en administrasie gefasiliteer en gestroomlyn kan word om doeltreffendheid te verhoog. Die vestiging, as 'n prioriteit, van die nodige meganismes vir toegang tot sulke databasisse vir die behoorlike koördinerings van funksies betreffende taalsake in Suid-Afrika.
 - Die transformering van die Staatstaaldiens en die Nasionale Terminologiesdiens tot 'n Nasionale Taaldiens (NTD) binne die DKKWT ten einde alle taalfasiliteringswerksaamhede in die Staatsdiens te moniteer en sodanige benodigde dienste te lewer.

 - As langtermynmaatreëls word voorsien:
 - Die vestiging van 'n stewige, landswye infrastruktuur van taaldienste wat 'n toereikende aantal goedopgeleide vertalers, tolke, terminoloë, terminografe, ensovoorts, insluit, asook spesialiswoordeboeke, woordeleste, en ander noodsaaklike materiaal, telefoniese en ander elektroniese fasiliteite vir mondelinge kommunikasie en 'n taalbedryf gerig op handel en ander interaksie met die Afrika-kontinent.
 - Die stel van haalbare tydraamwerke vir die ontwikkeling en gebruik van die gemarginaliseerde en benadeelde tale in hoërstatusfunksies.

Die *National Language Policy Framework* (NLPF) (DAC 2002) vloei voort uit die LANGTAG-verslag. Beginsels rondom terminologieontwikkeling en -bestuur word uitdruklik in afdelings 3.3, 3.6 en 4.9 genoem. Meer konkrete meganismes word egter voorgestel in die afdeling "Implementation Plan: NLPF". Die NLPF (vgl. DAC 2002: 18-20) voorsien die volgende:

- dat die implementering van die NLPF sal uitloop op 'n hoë vraag na vertaling en redigering, en die ontwikkeling van terminologie op alle terreine sal daarom kritiek wees, wat die samewerking van alle belanghebbendes, taaleenhede en taalliggame (bv. PanSAT-substrukture soos Nasionale Leksikografiese Eenhede, Nasionale Taalliggame en Provinsiale Taalkomitees) sal vereis,

- dat die Terminologiekoördineringsafdeling (TKA) van die NTD aktiewe samewerking deur bemiddeling van die Nasionale Taalforum sal bekom,

- dat 'n gerekenariseerde Nasionale Termbank deur die DKK gevestig sal word om toegang tot en verspreiding van terminologie in al die amptelike tale te fasiliteer, en
- dat die Termbank via die Internet toeganklik sal wees.

Hierdie maatreëls word beoog vir 2003 en daarna (DAC 2002: 28 — Summary: Overall Implementation Plan for NLPF).

Die vermelde dokumente word primêr gerig op terminologiebestuur binne die NTD, die instansie wat histories daarmee belas was. Daar bestaan egter in Suid-Afrika ook 'n tweede nasionale instansie wat vir terminologie verantwoordelik is, naamlik die Pan-Suid-Afrikaanse Taalraad (PanSAT). Hierdie instansie word grondwetlik vermeld as die liggaam primêr aanspreeklik vir taalontwikkeling.

PanSAT funksioneer as 'n onafhanklike statutêre liggaam wat gestig is in terme van die Pan-Suid-Afrikaanse Taalraad Wet (Wet nr. 59 van 1995) om veeltaligheid te bevorder en spesifiek om die voorheen ingekorte tale te ontwikkel. Ingevolge sy wet (soos geamendeer deur Wet 10 van 1999) het PanSAT drie tipes strukture geskep vir al die amptelike tale, die Khoe- en Santale, Suid-Afrikaanse Gebaretaal asook die erfenistale, te wete Provinsiale Taalkomitees, Nasionale Taalliggame en Nasionale Leksikografiese Eenhede:

- Die rol van die Provinsiale Taalkomitees (PTK's) in elke provinsie is om na die bevordering en ontwikkeling van die verskillende amptelike tale binne daardie provinsie om te sien. Hulle dien as strukture wat die Raad adviseer oor enige taalaangeleenthede. Hulle werksterrein is meer gerig op taalstatusbeplanning en prestigebepanning, maar ook taalverwerwingsbeplanning (vgl. Kaplan en Baldauf 2003).
- Die Nasionale Taalliggame (NTL'e) adviseer die Raad oor enige spesifieke taal, maar in die besonder oor taalontwikkeling vir daardie taal, wat insluit taalstandaardisering, maar ook spesifiek terminologie-ontwikkeling vir die amptelike tale.
- Die rol van die Nasionale Leksikografiese Eenhede (NLE's) is om woordeboeke vir al die tale saam te stel. Hulle funksie by taalontwikkeling lê dus veral by leksikale modernisering. PanSAT beskik egter nie oor 'n terminologiesentrum nie en maak dus ten opsigte van terminologie-aangeleenthede primêr staat op advies wat deur die NTL'e voorsien word, asook opdragnavorsing daaroor (PanSALB Annual Report 2004/2005).

Taalbeleidsmatig word daar gevolglik gesorg vir taalontwikkeling vir die agtergestelde tale van Suid-Afrika. Beleid wat tot hiertoe opgestel is, maak voorsiening vir twee nasionale rolspelers as terminologiebestuursagente. Die basis bestaan derhalwe vir die toepassing van die taalbeplanningsgeoriënteerde werkmetode by terminologiebestuur.

Vir hierdie werkmodes om te slaag, is dit noodsaaklik dat terme geskep, versamel en gesistematiseer sal word ten einde taalgebruikers in staat te stel om die geormerkte tale te kan gebruik. Uit 'n baie voorlopige ondersoek na die werkmodes wat tans gevolg word, blyk dit dat so 'n proses inderdaad op gang is, maar afgestem op die unieke Suid-Afrikaanse omstandighede. Daar sal saaklik stilgestaan word by van die terminologie-aktiwiteite wat deur die gemelde rolspelers gehanteer word.

4.3.1 Die Terminologiekoördineringsafdeling van die NTD

Die NTD se Terminologiekoördineringsafdeling het sedert 1994 verskeie veranderinge ondergaan om aan te pas by die nuwe uitdaging om terminologie-uitsette in al die amptelike tale beskikbaar te maak. Uit die tipe meertalige termlyste wat sedertdien saamgestel is, soos byvoorbeeld *Inheemse soogdiere* (1996), *Basiese gesondheidsterme* (1997) en *MIV/Vigs-bewustheidsveldtog-terme* (1998), kan afgelei word dat daar 'n registerverandering by terminologiewerk ingetree het. Die doelwit word meer prakties en basies van aard en dit wil voorkom asof daar nie slegs in spesifieke vakgebiede op die formele register gekonsentreer word nie, maar eerder op alle domeine van kennis en menslike aktiwiteite waar daar 'n behoefte bestaan om die gebruikers van die tale van doeltreffende kommunikasie te voorsien. Die tipiese termlyste wat tans saamgestel word, bevat Engelse terme as die bronsterme. Al die bronsterme word voorsien van eenvoudige verduidelikings of definisies in maklik verstaanbare Engels met ekwivalente in Afrikaans, Zulu, Xhosa, Swati, Ndebele, Tswana, Noord-Sotho, Suid-Sotho, Venda en Tsonga of 'n kombinasie van hierdie tale as doeltale (Alberts 2000: 9-10).

In Maart 2000 het die Minister van Onderwys 'n versoek aan dr. Ben Ngubane, destydse Minister van Kuns, Kultuur, Wetenskap en Tegnologie gerig dat die Terminologiekoördineringsafdeling basiese terminologielyste saamstel om leerders te help wanneer hulle in graad 5 vanaf moedertaalonderrig na Engels oorskakel. Die fokus van hierdie terminologielyste is op die uitbreiding en bevordering van terminologiedata in die amptelike tale om as ondersteunende hulpmiddels by opleiding te dien.

Die aktiwiteitsterreine wat as prioriteitsgebiede geïdentifiseer word, is Wiskunde (Mathematics), Natuurwetenskap (Natural Science) en Tegnologie (Technology). Die projekte vir Wiskunde en Natuurwetenskap is al afgehandel en aan die NTL'e vir goedkeuring en verifiëring voorgelê. Die projekte vir die Sosiale en Ekonomiese Wetenskappe is reeds begin, is goed op dreef en word ten opsigte van strukturele bystand ondersteun deur PanSAT se NTL'e (PanSALB Annual Report 2004/2005).

Wanneer die werksaamhede van die Terminologiekoördineringsafdeling in oënskou geneem word, blyk dit dat die fokus rondom terminologiebeplanning meer prakties en basies is as wat die geval tydens die vorige taalbestel was toe 'n linguïsties georiënteerde saam met 'n vertalingsgeoriënteerde bena-

dering gevolg is. Hierdie fokus word ondersteun deur die feit dat basiese terminologielyste in die amptelike tale saamgestel word en dat terminologie vir aktiwiteitsterreine by onderwys vir junior leerders tans 'n prioriteitsgebied vorm.

4.3.2 PanSAT en sy NTL'e

Wat betref die ontwikkeling van terminologie werk PanSAT en sy NTL'e saam met die Suid-Afrikaanse Standaardiseringskomitee van ISO/TC 37 (StanSA TC 37) by die Suid-Afrikaanse Buro vir Standaarde (SABS). Hiervoor is Tegniese Komitees vir Terminologiewerk tot stand gebring. PanSAT werk ook saam met die nuwe Departement Kuns en Kultuur (DKK, plaasvervanger van die DKKWT) en die Departement van Onderwys om terminologie te ontwikkel in leerareas wat betrekking het op Wetenskap en Tegnologie en om strukturele bystand in hierdie verband te verleen. Dr. M. Alberts, die PanSAT-bestuurder vir Standaardisasie en Terminologie, is ook deurlopend besig met terminologie-opleiding by verskeie instansies. In haar opleiding gee sy aandag aan al drie werkmodes van terminologiebestuur. PanSAT is verder betrokke by spesifieke navorsing ten opsigte van terminologieontwikkeling (PanSALB Annual Report 2004/2005).

Die NTL'e se Tegniese Komitees vir Terminologiewerk verifieer terme en verleen goedkeuring aan wat onder meer deur die NTD en ander instansies of private inisiatiewe gedoen word. Die NTL'e se Tegniese Komitees: Standaardisering het tussen 2005 en 2006 aan die hersiening van die spelreëls van die Afrikaanse gewerk. Deel daarvan is om ook na die terminologielyste te kyk (bv. woorde wat maklik verkeerd gespeld word of nuutskeppings is). Selfs die Nasionale Taallegging vir Suid-Afrikaanse Gebaretaal is tans besig met terminologieontwikkeling. Dit maak video-opnames van nuwe terme om dié terme te probeer standaardiseer (PanSALB Annual Reports 2003/2004 en 2004/2005).

Koördinering tussen PanSAT en die NTD met betrekking tot terminologieontwikkeling is uiters noodsaaklik aangesien duplisering van take rondom hierdie kwessie maklik kan voorkom. Alhoewel die situasie nog nie heeltemal ideaal is nie, blyk dit dat daar op die oomblik wel koördinering plaasvind en dat die meeste projekte wat deur die NTD, die Terminologiekoördineringsafdeling en ander instansies of private inisiatiewe gedoen word, aan die NTL'e voorgelê word vir verifiëring, bekragtiging en die uiteindelijke produksie van terminografiese produkte.

5. Slotsom

In hierdie artikel is drie heersende terminologiese werkmodes by terminologiebeplanning, -koördinering en -bestuur oorweeg, te wete die linguisties geïntegreerde werkmodes, die vertalingsgeïntegreerde werkmodes en die sogenaamde taalbeplanningmetode. Die implikasies van die verskillende werk-

metodes vir terminologieontwikkeling en -bestuur in Suid-Afrika is ondersoek. Daar is bevind dat die taalbeplanningsgeoriënteerde werkmodes die aangewese metode vir die huidige Suid-Afrikaanse situasie is. Hierdie metode word meestal aangewend in meertalige situasies waar 'n voormalige koloniale taal dominant is en waar 'n inheemse taal (tale) as amptelike taal (tale) ontwikkel moet word. Terminologie word uit die dominante taal ingevoer na die inheemse taal(tale) en gepaste ekwivalente word na gelang van behoefte ontwikkel. Die sukses van hierdie metode berus egter op die mate waartoe beleids- en ander maatreëls die terminologieproses ondersteun.

Die taalbeplanningsgeoriënteerde werkmodes by terminologiebestuur blyk gepas te wees vir die Suid-Afrikaanse situasie. Die wetgewende raamwerk (verampteliking van elf tale) is geskep vir die implementering van die taalbeplanningsgeoriënteerde werkmodes. Engels word reeds in die meeste vakgebiede en domeine as kommunikasiemiddel gebruik. Die taal kan derhalwe volgens die taalbeplanningsgeoriënteerde werkmodes aangewend word vir die onttrekking van terme en die saamstel van definisies. Die ander tien amptelike tale kan as doeltale funksioneer waarin ekwivalente verskaf moet word. Die suksesvolle toepassing van die taalbeplanningsgeoriënteerde werkmodes kan tot gevolg hê dat terme versamel, gesistematiseer en voorberei word vir gebruik op verskillende vlakke en registers. Gebruikers sal dan in staat wees om makliker en doelmatiger in die amptelike tale te kommunikeer wat die status van hierdie tale sal verhoog. Terminologie-inligting in die onderontwikkelde tale sal beskikbaar gestel word wat hierdie tale vir hoëfunksietaalgebruik geskik sal maak.

Daar moet dus tot die gevolgtrekking gekom word dat die linguisties georiënteerde saam met die vertalingsgeoriënteerde metode soos wat dit tot voor 1994 by terminologiebestuur in Suid-Afrika toegepas is, uiteraard nie meer geskik is nie. As gevolg van 'n nuwe taalklousule wat die verampteliking van 'n bykomende nege inheemse tale vereis, moet Suid-Afrika nou wat terminologiebestuur betref, as 't ware 'n paradigmaskuif maak.

Daar behoort verder vasgestel te word of daar voldoende vordering is met die oorgang na so 'n nuwe stelsel waar die taalbeplanningsgeoriënteerde metode gebruik word en of al die nodige vereistes vir die suksesvolle toepassing daarvan nagekom word. Ook behoort ondersoek gedoen te word na die implikasies wat die oorgang na hierdie nuwe stelsel het vir die voortgesette ontwikkeling van 'n taal soos Afrikaans wat nie as 'n onderontwikkelde taal naas die inheemse tale beskou kan word nie en wat besonder gebaat het by die linguisties georiënteerde saam met die vertalingsgeoriënteerde metode by terminologiebestuur onder die vorige taalbestel.

Notas

1. Die term "Afrikatale" bly linguisties gesproke problematies, maar is binne die Suid-Afrikaanse konteks steeds moeilik vervangbaar met die korrekter term "Bantoetale". By gebrek aan 'n

beter oplossing word die term "Afrikatale" hier in die populistiese sin van die woord as ekwivalent van die ewe problematiese begrip "swart tale" (*sic*) gebruik.

2. Language for Special Purposes (LSP), language used in a subjectfield and characterized by the use of specific linguistic means of expression that include subject-specific terminology and phraseology and also may cover stylistic or syntactic features (ISO 1087-1: 6).

Bibliografie

- Alberts, M.** 2000. *Audit on South African Terminology Projects. July 2000.* Ongepubliseerde dokument.
- Auger, P.** 1988. La terminologie au Québec et dans le monde, de la naissance à la maturité. Gaumont, J.-C. et al. (Reds.). 1988. *L'ère nouvelle de la terminologie. Actes du sixième colloque OLF-STQ de terminologie, Montréal du 27 au 29 novembre 1985: 27-59.* Montreal: L'Office.
- Cabré, M.T.** 1999. *Terminology: Theory, Methods and Applications.* Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Cluver, A.D. de V. en C.J. Scheffer (Reds.).** 1984. *RGN-ondersoek na vaktaalaangeleenthede in Suidelike Afrika: Verslag van die ondersoek na organisatoriese aangeleenthede op vaktaalgebied. Term-1.* Pretoria: RGN.
- Department of Arts and Culture (DAC).** 2002. *National Language Policy Framework. Final Draft. 13 November 2002.* Pretoria: Department of Arts and Culture.
- Department of Arts, Culture, Science and Technology (DACST).** 1996. *Overview, Recommendations and Executive Summary from the Final Report of the Language Plan Task Group (LANGTAG).* Pretoria: Department of Arts, Culture, Science and Technology.
- Ferguson, C.A.** 1972. Diglossia. Giglioli, P.P. (Ed.). 1972. *Language and Social Context: Selected Readings: 72.* Harmondsworth: Penguin Education.
- Grondwet.** 1996. Die Republiek van Suid-Afrika. Wet 108 van 1996.
- ISO 1087:** 1990. *Terminology — Vocabulary.* International Organization for Standardization.
- ISO 10241:** 1992. *International Terminology Standards — Preparation and Layout.* International Organization for Standardization.
- ISO 12616:** 2002. *Translation-oriented Terminography.* International Organization for Standardization.
- Kaplan, R.B. en R.B. Baldauf Jr.** 2003. Language Planning in Perspective. Trends from Diversity. Kaplan, R.B. en R.B. Baldauf Jr. (Eds). 2003. *Language and Language-in-Education Planning in the Pacific Basin: 201-226.* Dordrecht/Boston/London: Kluwer Academic Publishers.
- Mouton, J.** 2005. Afrikaans as wetenskaptaal in Suid-Afrika. *Tydskrif vir Geesteswetenskappe* 45(3): 370-385.
- Pan South African Board Act.** 1995. Act Nr. 59 of 1995.
- Pan South African Language Board Amendment Act.** 1999. Act Nr. 10 of 1999.
- Pan South African Language Board.** 2003/2004. *Annual Report.*
- Pan South African Language Board.** 2004/2005. *Annual Report.*
- Sager, J.C.** 1990. *A Practical Course in Terminology Processing.* Amsterdam: John Benjamins Publishing Company.
- Wright, S.E. en G. Budin.** 1997. Introduction. Wright, S.E. en G. Budin (Reds.). 1997: 1-12.
- Wright, S.E. en G. Budin (Reds.).** 1997. *Handbook of Terminology Management.* Amsterdam: John Benjamins Company.
- Wüster, E.** 1979. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie.* Abhandlungen zur Sprache und Literatur 20. Wene/New York: Springer.

A Transtextual Approach to Lexicographic Functions*

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

Abstract: The development of theoretical lexicography clearly indicates a movement towards a more general recognition of the role of lexicographic functions. A consistent application of lexicographic functions has a major influence on the contents of dictionary articles but also on the data distribution and even the typological classification. In the discussion of lexicographic functions, the emphasis has primarily been on the central list of dictionaries. This article proposes an approach where outer texts are also selected and compiled in order to make an active contribution to achieving the lexicographic functions of a given dictionary. A distinction is made between function-adhering and non-function-adhering outer texts. The function-adhering outer texts can further be divided into integrated and non-integrated function-adhering outer texts. With reference to the data distribution programme of a dictionary, it is suggested that in the planning of every dictionary, provision should be made for a function allocation programme. This will ensure that the identified functions of a given dictionary will be realised in both the central list and the outer texts.

Keywords: BACK MATTER TEXTS, CENTRAL LIST, COGNITIVE FUNCTION, DATA DISTRIBUTION, FRAME STRUCTURE, FRONT MATTER TEXTS, FUNCTION ALLOCATION PROGRAMME, FUNCTION-ADHERING OUTER TEXTS, FUNCTIONAL TEXT PARTS, INTEGRATED FUNCTION-ADHERING OUTER TEXTS, LEXICOGRAPHIC FUNCTIONS, NON-FUNCTION-ADHERING OUTER TEXTS, NON-INTEGRATED FUNCTION-ADHERING OUTER TEXTS, OUTER TEXTS, TEXT PRODUCTION, TEXT RECEPTION, TRANSTEXTUAL APPROACH

Opsomming: 'n Transtekstuele benadering tot leksikografiese funksies.

Die ontwikkeling van die teoretiese leksikografie toon 'n duidelike beweging in die rigting van 'n groter erkenning aan die rol van leksikografiese funksies. 'n Konsekwente toepassing van leksikografiese funksies het 'n ingrypende invloed op die inhoud van woordeboekartikels maar ook op die dataverspreiding en selfs op die tipologiese klassifikasie. Tot dusver was die klem in die gesprek oor leksikografiese funksies veral op die sentrale teks van woordeboeke. Hierdie artikel bepleit 'n benadering waar die buitetekste ook gekies en saamgestel moet word om aktief by te dra tot die bereiking van 'n bepaalde woordeboek se leksikografiese funksies. 'n Onderskeid word gemaak tussen funksienakomende en niefunksienakomende buitetekste. Die funksienakomende buitetekste kan op hulle beurt weer onderverdeel word in geïntegreerde en niegeïntegreerde funksienakomende tekste. Na aanleiding van die dataverspreidingsprogram van woordeboeke word voorgestel

* This article was presented as a paper at the Thirteenth International Symposium on Lexicography at the University of Copenhagen, Copenhagen, Denmark 3–5 May 2007.

dat daar in die beplanning van elke woordeboek ook voorsiening gemaak word vir 'n funksietoekennende program. Dit sal verseker dat die geïdentifiseerde funksies van 'n bepaalde woordeboek in sowel die sentrale teks as die buitetekste verwerklik word.

Sleutelwoorde: AGTERTEKSTE, BUITETEKSTE, DATAVERSPREIDING, FUNKSIENAKOMENDE BUITETEKSTE, FUNKSIETOEKENNENDE PROGRAM, FUNKSIONELE TEKSDELE, GEÏNTEGREERDE FUNKSIENAKOMENDE BUITETEKSTE, KOGNITIEWE FUNKSIE, LEKSIKOGRAFIESE FUNKSIES, NIEFUNKSIENAKOMENDE BUITETEKSTE, NIEGEÏNTEGREERDE FUNKSIENAKOMENDE BUITETEKSTE, RAAMSTRUKTUUR, SENTRALE TEKS, TEKSPRODUKSIE, TEKSRESEPSIE, TRANSTEKSTUELE BENADERING, VOORTEKSTE

1. Introductory remarks

The development of theoretical lexicography shows a number of significant trends. The lexicographic practice developed in a pre-theoretical era and theoretical lexicography has to be regarded as a relative late-comer, complementing practical lexicography. One of the important early phases in the development of theoretical lexicography was heralded by the publication of Zgusta (1971). This phase of theoretical lexicography had a strong linguistic bias with the focus primarily on the linguistic contents of dictionaries. This linguistic approach had a wide-ranging influence, among others also on an aspect like typological classification. At this stage, theoretical lexicography was primarily practised by linguists and taught in linguistic or language departments. Lexicography was seen as a subdiscipline of linguistics.

In the course of time it became clear that language is the object of linguistics and dictionaries are the objects of lexicography. Consequently lexicography was no longer seen as subdiscipline of linguistics but as a discipline in its own right (cf. Wiegand 1998: 33, Tarp 2000: 192, Hartmann and James 1998, Bergenholtz and Tarp 2005). Although the linguistic contents of dictionaries remained an important aspect of theoretical lexicography, there developed a keen interest in the packaging of this linguistic data and a shift in emphasis towards the structure of dictionaries. The user-driven approach led to yet another phase in the development of theoretical lexicography, i.e. lexicographic functions. This development in theoretical lexicography has led to numerous changes in the way lexicographers speak about dictionaries and, hopefully, in the way people compile dictionaries. Unfortunately the full consequence of these changes, and this applies to the emphasis on dictionary structures as well as the subsequent emphasis on lexicographic functions, has not yet been realised — neither in theoretical lexicography nor in lexicographic practice. One such example where the implications of lexicographic functions have not yet been applied in a consistent way is the typological classification of dictionaries where the traditional typological models are still largely utilised. Naturally, many of these classifications are still valid and should remain in use. As an example: there is not really a classificatory substitute for etymological dictionary and even the distinction

between diachronic and synchronic dictionaries, albeit a linguistically-driven distinction, still has a definite value in modern lexicography. But a functional approach to dictionary typology should at least see the envisaged functions of a dictionary being displayed as a subtitle or another kind of entry on the cover and title page of a printed dictionary or on an opening screen of an electronic dictionary. In doing this, lexicographers do not have to confront their users with technical terms like *text production*, *text reception* or *cognitive function*, but a traditional title may be supplemented by an entry like: "helping with the writing/understanding of texts". The implications of a functional approach for typological classification is a matter for a different article. The idea has been discussed briefly in Hartmann (1987), Tarp (2002) and, in slightly more detail, in Gouws (to appear).

An illustration of where developments in lexicographic structures have not yet been applied in an optimal way in the lexicographic practice regards the frame structure. Although dictionaries more often than not employ front and back matter texts as venues for a more comprehensive data distribution, these outer texts are too often selected on either an arbitrary or a traditional basis; traditional in the sense that it exclusively relies on the pattern followed in other dictionaries without any innovative approaches in terms of new types of outer texts or a new way of presenting data in these texts. Too seldom they are employed to enhance the realisation of the lexicographic functions of a specific dictionary.

In this regard dictionary users should be introduced to the use of functions as more than merely nice-to-haves or a presentation of additional data not as relevant to the dictionary as the data included in the central list. Lexicographers should employ their own dictionaries to familiarise users with the functions of the dictionary, as realised in both the central list and the outer texts. A front matter text like the user's guide should be employed to explicate the nature of e.g. dictionary structures and lexicographic functions. Nielsen (2005: 137) states the twofold purpose of the user guide of a dictionary as follows:

First of all, it introduces the user to the structure and contents of the dictionary so that he may have the greatest possible benefit of its data. Secondly, it explains the fastest and easiest way to the data and hence the information sought.

This needs to be seen as criticism against the way in which many lexicographers do not use the user guides of their dictionaries. Although Nielsen does not mention anything about lexicographic functions in this paragraph, he later criticises the lack of a discussion of lexicographic functions in the user's guide of the majority of dictionaries when saying that "user's guides whose texts are explicitly based on dictionary functions are few and far between". He also makes a plea for an explicit link between the functions of a dictionary and the information in its user's guide (Nielsen 2005: 138).

The approach of lexicographic functions needs a much more aggressive application in the outer texts of dictionaries, especially when considering Tarp

(2000: 198) who maintains that:

The functions are the very essence of lexicography. The functions constitute the leading principle of all dictionaries. Everything in a dictionary is to a greater or lesser extent influenced by its respective functions. Neither the contents nor the form of a dictionary can be conceived without taking the functions into account. It is possible to outline an integrated model of a lexicographic reference work on the basis of the functions.

In the remainder of this article the emphasis will be on the way in which a consistently applied functional approach could influence the theory regarding outer texts as well as the selection of outer texts and the presentation of data in these texts. However, first some comments will be made regarding the status of outer texts, especially the user's guide text, as targets of dictionary consultation procedures.

2. Realising the importance of outer texts

When speaking about outer texts metalexicographers emphasise their importance whilst practical lexicographers often maintain that these texts, especially the users' guidelines text presented in the front matter section of dictionaries, are rarely consulted by dictionary users. Looking at the aims of dictionaries of languages used in Central Africa, Busane (1990: 22) states that by looking at the prefaces and introductions of dictionaries one can obtain information about the audience at whom the dictionary had been directed. He goes further by referring to Haugen (1985) who used this method to determine the goals envisaged by Scandinavian lexicographers and who indicated that by reading numerous introductions to dictionaries one can learn what the purposes were "that animated lexicographers to undertake their often tedious and always underpaid tasks". Busane gives examples from front matter texts identifying the intended target users. Even though outer texts may be compiled in such a way that they have a user-friendly nature, a dictionary culture is needed to ensure the use of outer texts. In this regard Busane (1990: 28) says:

Many introductory pages (are) usually allocated to grammatical sketches of the language concerned without the knowledge of which it is deemed hazardous to use the dictionary successfully. We believe, however, that these sections and introductory explanations are not sufficient provisions for a user-friendly product. Dictionary users are known to allocate little time to the study of these prefatory matters.

The user also comes to the fore in Herberg (1989: 751) who discusses problems regarding the prefaces of dictionaries and who indicates, also in this regard, the importance to distinguish between different user groups, i.e. laypersons and scientists, so that the dictionary can respond to the demands of both these user groups. He continues that the prefaces of dictionaries should be compiled as

purposeful and user-oriented as possible. According to him, the nature of the preface of a dictionary depends largely on the type of dictionary. Because dictionary types should be compiled for specific users, one can deduce that the target user group, their needs and reference skills should also play a decisive role in the planning and compilation of outer texts.

When Herberg's approach (that the prefaces of dictionaries should be compiled as purposeful and user-oriented as possible) is followed, and in spite of the scepticism of Busane shared by many others, the status of outer texts as venues of dictionary consultation may not be eschewed. In this regard promising confirmation comes from the Danish phraseological dictionary *Ordbogen over Faste Vendinger* and the Danish music dictionary *Den Danske Musikordbogen*, both internet dictionaries. An analysis of the log files of these two dictionaries indicates unequivocally that users do consult the outer texts. Over a brief period of time, an average of at least five percent of the daily users of the phraseological dictionary consulted the various outer texts. Similar results were obtained from the use of the music dictionary.[†] Interestingly enough, the log file indicated that the preface of the phraseological dictionary has more hits than the user's guide. This has implications for lexicographers believing that the preface does not have a role to play in a dictionary. Bearing in mind that the number of consultations of outer texts is given as a percentage of the total daily consultation count, one can expect that the actual percentage of users who consult the outer texts, albeit not on a daily basis, could be considerably higher. The same user will consult the dictionary more than once, but will not necessarily use an outer text every time. These results put the responsibility on the shoulders of the lexicographer to utilise a frame structure and to include the types of outer texts needed by the target users of the envisaged dictionary. The lexicographer should not only provide outer texts with satisfactory contents, but should also ensure that the functions of the dictionary prevail in these outer texts.

3. Changes in the theoretical approach to outer texts

Wiegand (1988: 844) argues that dictionaries are utility products and that they have a genuine purpose. Bergenholtz, Tarp and Wiegand (1999: 1776) take the notion of the genuine purpose of dictionaries further by saying that the genuine purpose of a dictionary consists in the possibility to retrieve information from the lexicographic data about the subject matter of the dictionary. The outer texts of a dictionary are also classified in terms of the genuine purpose, distinguishing between integrated and non-integrated outer texts. Outer texts from which it is possible to retrieve information that helps to fulfil the genuine purpose of a dictionary are called *integrated outer texts*. This important distinction between integrated and non-integrated outer texts, a distinction based on the role of the outer texts regarding the fulfilment of the genuine purpose of a dictionary, has not yet found a parallel in the function-based theoretical ap-

proach. Negotiating the point of view of Tarp (2002: 198) that everything in a dictionary is to a greater or lesser extent influenced by the functions of the dictionary, it is somewhat disturbing that the theoretical discussions of lexicographic functions have not yet given due acknowledgement to the extent of the nature and scope of the implications of lexicographic functions for other texts in the dictionary besides the central list. Outer texts should also be selected and compiled in order to achieve the function(s) of a given dictionary. A distinction that needs to be made is that between a different kind of integrated and non-integrated outer texts, i.e. outer texts which are integrated or not integrated into the function(s) of a dictionary (cf. Gouws and Steyn 2005).

Within the category of outer texts of a given dictionary, a distinction should be made between those texts that do and those that do not enhance the information transfer in the dictionary. Here the notion of information transfer refers to texts that are consulted by users as part of the overall dictionary consultation process, and not those texts that, relevant as they may be, contain data that does not constitute a target of a typical dictionary consultation procedure, e.g. the imprint or a text in which the lexicographer expresses thanks to sponsors, colleagues, family, etc.

Wiegand (1989: 425) distinguishes between functional and non-functional text segments, indicating that the difference lies in the genuine purpose of the segment. A functional text segment is a part of a dictionary article, consisting of a form and at least one genuine purpose. Here the word *functional* should not be confused with lexicographic functions. Hausmann and Wiegand (1989: 330) indicate that the front matter of a dictionary as a whole is not a functional part of the dictionary, but rather an arbitrary set of functional text parts. They refer to the different front and back matter texts as *functional text parts*. Yet again, *functional* should not be seen as referring to a lexicographic function. According to them, front matter texts like the title, the table of contents and the imprint are all functional text parts. Following Wiegand's definition of functional text segments, one could go further and take this notion from the level of the text segment to the level of the text part by arguing that a functional text part, here an outer text, would be a text part with a form and at least one genuine purpose. This genuine purpose does not necessarily have to agree with the genuine purpose of the dictionary as a whole.

Within a lexicographic function approach, a distinction needs to be made between Wiegand's functional outer texts, i.e. outer texts that have a genuine purpose, and function-adhering outer texts, i.e. outer texts by means of which a given lexicographic function can be achieved. Within any given dictionary, the function-adhering outer texts should be distinguished from non-function-adhering outer texts, i.e. the imprint, the list of staff and contributors, etc. Function-adhering outer texts can be divided into integrated function-adhering outer texts and non-integrated function-adhering outer texts. The integrated function-adhering outer texts are integrated into the lexicographic functions of the central list of the given dictionary. This implies that the functions identified

for the specific dictionary are not restricted to the central list, but go beyond the borders of this text to allow a transtextual function approach. The distinction between integrated and non-integrated function-adhering outer texts allows the lexicographer to include outer texts that assist users in achieving a well-identified lexicographic function, albeit that this function has not been identified as having to prevail in the central list of the dictionary. Such an approach adds a new dimension to the notion of polyfunctional dictionaries, i.e. that all the functions do not have to be achieved in the central list of such a dictionary and that the central list is not the only venue for the realisation of functions.

4. Putting new developments in lexicographic theory to practice

In current lexicographic practice, the awareness level of lexicographic functions in the outer texts of dictionaries is not satisfactory. Too often it still seems as if the inclusion of outer texts has been done on a random basis without negotiating the genuine purpose of a given dictionary. Tarp (2002: 16) says:

According to the functional theory of lexicography, no data whatsoever should be included in any dictionary if it cannot be argued on the basis of its respective functions.

In some dictionaries, a functional and even a transtextual functional approach prevails without the user being convinced that a lexicographic function approach had been planned for the dictionary. A more deliberate application of lexicographic functions in the outer texts and an unambiguous selection of these texts and their contents in terms of the identified lexicographic functions will enhance the information transfer in a dictionary.

Albeit that a transtextual functional approach does not always prevail, outer texts, especially the user's guide, have been put to use to explicate the functions of a given dictionary and in so doing have accommodated the results of metalexicographic research; cf. the way in which the prevailing functions of the dictionary are introduced in the internet version of the *English Dictionary of Accounting*:

A dictionary is an aid, a tool. Like other tools, some dictionaries have one function and others have several functions. The *English Dictionary of Accounting* has several functions. ...

The tool called a dictionary can be used in connection with **specific communication-related problems** or in connection with **a wish to learn more**. ...

Communication-related information serves to help you solve problems that may occur when you **write** or when you **read** English accounting texts.

Help to write English texts

When **writing** accounting texts in English you may not be sure about the correct spelling of a word. This can be found in the presentation of the headword. ...

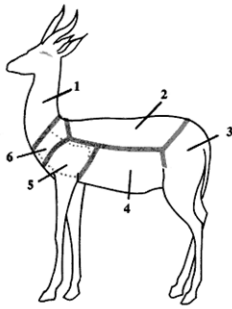
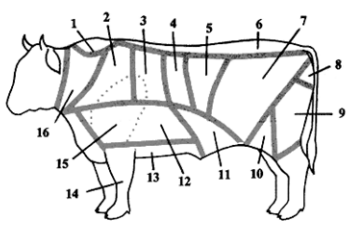
Help when reading English texts

When you read an English accounting text, you may be in doubt as to the correct meaning of a term. ...

Help to acquire knowledge about accounting matters

The functions described above serve as help to solve problems that are related to specific types of communicative acts. In addition, this Dictionary can also be used if you want to learn more about something, i.e. in connection with knowledge-oriented problems.

Introducing the lexicographic functions of a given dictionary in its user's guide can be regarded as a first step in the utilisation of outer texts to enhance the realisation of lexicographic functions. A next step would be the deliberate choice of outer texts on the basis of these texts adhering to the lexicographic functions identified for the given dictionary and the announcement of these functions, and the transtextual functions in the user's guide of the dictionary.

Cuts of meat		Vleissnitte	
<p>Venison (Game)</p> <p>neck 1 loin and saddle 2 leg 3 flank 4 shoulder 5 breast 6</p> <p>venison chops venison steak jugged venison roast venison</p> <p>venison pie venison stew jointed guinea fowl ripen game game biltong</p>	<p>Wildsvleis</p> <p>nek 1 lende en saal 2 'boud 3 lies 4 blad 5 bors 6</p> <p>wildstjop wildskyf fyngestooftde wildsvleis gebraaide wildsvleis/ wildsbrand wildspastei wildsbredie ligedeelde tarentaal wildsvleis laat ryp word wildsbiltong</p>		
	<p>Beef</p> <p>hump 1 chuck 2 prime rib 3 wing rib 4 sirloin 5 fillet 6 rump 7 aitchbone 8 silverside 9 thick flank 10 thin flank 11 flat rib 12 brisket 13 shin 14 bolo 15 neck 16</p> <p>mince rump steak topside</p>	<p>Beesvleis</p> <p>blad 1 dikrib 2 primarib 3 voorrib 4 beeslende 5 beeshaas, beesfilet 6 kruisstuk 7 stertstuk, ysbeen 8 dystuk 9 diklies 10 dunlies 11 platrib 12 beesbors 13 skenkel 14 bolo 15 nek 16</p> <p>maalvleis, gemaalde vleis (fynvleis) kruisskyf binneboud</p>	

The preface of the Afrikaans–English/English–Afrikaans bilingual dictionary *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary* gives no indication of the envisaged target user or the lexicographic functions of the dictionary although it explicitly refers to the assistance users get in "using the right word" and "understanding the other language", i.e. text production and

text reception respectively. When one looks at the central list, it soon becomes clear that this dictionary primarily has a communication function with the emphasis especially strong on text reception and to a lesser degree also on text production. The back matter section of this dictionary contains a range of outer texts (cf. Gouws 2004). Some of these texts maintain the communication functions of the central list and help to establish a transtextual application of lexicographic functions. Some other back matter texts also adhere to a lexicographic function albeit that they clearly deviate from the functions typically prevailing in the central list of the dictionary. A back matter text presenting the different cuts of meat for beef, pork and mutton, along with pictorial illustrations to show where each cut is found on the relevant animal, typically has a cognitive function. The same applies to a back matter text presenting a map of the world with a wide-ranging selection of oceans, continents, countries, capital cities, etc. indicated on this map. These outer texts with their cognitive function are not integrated into the functions of the central list of the dictionary, but being function-adhering outer texts, they increase the polyfunctional value of the dictionary.

Awareness of the data distribution structure (cf. Bergenholtz, Tarp and Wiegand 1998), the frame structure (cf. Kammerer and Wiegand 1998) and the notion of a dictionary being a carrier of text types (cf. Wiegand 1996) has given lexicographers the freedom to go beyond the central list in the presentation of data. With regard to the structure and contents of dictionaries, the required response has been obtained, but not yet with regard to lexicographic functions. A stronger interaction in terms of lexicographic functions between outer texts and central list is needed. In this regard lexicographers should, as part of the dictionary conceptualisation plan, negotiate the possibility of a function allocation programme, similar to the data distribution structure, to ensure that all the functions identified for a specific dictionary will be realised and that both outer texts and central list participate in the realisation of these functions. A lack of a well-planned function allocation programme too often results in dictionaries with text production as function to have back matter sections dominated by texts with a text reception or cognitive function. These texts are important, but represent a secondary function of the dictionary. Outer texts are not only there for the secondary functions, but should also be employed for a realisation of the primary function(s). For a text production function, back matter texts like *Writing a letter*, *Forms of address*, *Collocations*, etc. can play an important role as function-adhering integrated outer texts. Once the primary functions have been achieved, the importance and value of function-adhering non-integrated outer texts may not be underestimated. In this regard the introduction of a cognitive function in outer texts in a dictionary where communication functions prevail in the central list can enhance the status of the dictionary as a response to the needs of a specific target user group. The function allocation programme should determine the textual venues where both primary and secondary functions should prevail.

5. In conclusion

As a carrier of text types, a dictionary displays different structures, contains different data types and can assist in achieving one or more lexicographic functions. In the same way in which metalexicographers argue in favour of the word book structure of a dictionary, focusing not only on the central list but also the outer texts in the front and back matter sections, metalexicographers who favour a lexicographic functions approach should work towards a situation where dictionaries are characterised by their transtextual function approach, with functions prevailing in different texts of the carrier of text types. Introducing the concepts of function-adhering outer texts and the distinction between integrated and non-integrated function-adhering outer texts, defined in terms of their relation to the identified functions of the given dictionary, this article favours a wide-ranging application of lexicographic functions across textual borders in order to enhance the success of typical dictionary consultation procedures. From a functional perspective, the typical comment of outer texts should be: the truth is out there.

Endnote

† Information obtained from Henning Bergenholtz — personal communication.

Literature

Dictionaries

- Bergenholtz, H. and V. Vrang. 2007. *Ordbogen over Faste Vendinger*. <http://www.idiomordbogen.dk>.
- Bergenholtz, I. 2006. *Den Danske Musikordbogen*. <http://musikordbogen.dk/musik/>.
- Grobbelaar, P. et al. (Eds.). 1987. *Reader's Digest Afrikaans-Engelse Woordeboek/English-Afrikaans Dictionary*. Cape Town: The Reader's Digest Association, South Africa (Pty) Ltd.
- Hartmann, R.R.K. and G. James. 1998. *Dictionary of Lexicography*. London: Routledge.
- Nielsen, S., L. Mourier and H. Bergenholtz. 2006. *English Dictionary of Accounting*. Aarhus: Aarhus School of Business.

Other literature

- Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen*. Frankfurt: Peter Lang.
- Bergenholtz, H. and S. Tarp. 2005. Wörterbuchfunktionen. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005: 11-25.
- Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999 *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Lan-*

- guages for Special Purposes. An International Handbook of Special-Language and Terminology Research: 1762-1832.* Berlin/New York: Walter de Gruyter.
- Busane, M.** 1990. Lexicography in Central Africa: The User Perspective with Special Reference to Zaïre. Hartmann, R.R.K. (Ed.). 1990. *Lexicography in Africa: 19-35.* Exeter Linguistic Studies 15. Exeter: University of Exeter Press.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H.** Forthcoming. Op pad na 'n nuwe woordeboektipologie. *Tydskrif vir Taalonderrig.*
- Gouws, R.H. and M. Steyn.** 2005. Integrated Outer Texts: A Transtextual Approach to Lexicographic Functions. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005: 127-136.
- Hartmann, R.R.K.** 1987. Four Perspectives on Dictionary Use: A Critical Review of Research Methods. Cowie, A.P. (Ed.). *The Dictionary and the Language Learner: 11-28.* Tübingen: Max Niemeyer.
- Haugen, E.** 1985. Lexicography and Language Planning. Jankowsky, K.R. (Ed.). *Scientific and Humanistic Dimensions of Language: 571-580.* Amsterdam: John Benjamins.
- 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: 328-360.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie.* Handbücher zur Sprach- und Kommunikationswissenschaft 5.1-5.3. Berlin: De Gruyter.
- Herberg, D.** 1989. Wörterbuchvorwörter. Hausmann, F.J. et al. (Eds.). 1989-1991: 750-754.
- Kammerer, M. and H.E. Wiegand.** 1998. Über die textuelle Rahmenstruktur von Wörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Nielsen, S.** 2005. User's Guides. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005: 137-146.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2002. Basic Elements of Lexicographic Theory/Éléments de base de la théorie lexicographique. Emejulu, J. du P. (Ed.). 2003. *Éléments de lexicographie gabonaise. Tome II: 7-35.* New York: Jimacs-Hillman.
- Wiegand, H.E.** 1988. Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischem Wissen. Munske, H.H. et al. (Eds.). *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern: 729-790.* Berlin: De Gruyter.
- Wiegand, H.E.** 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989-1991: 409- 462.
- Wiegand, H.E.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger lexikographischen Kolloquium: 1-82.* Lexicographica. Series Maior 76. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* Berlin: De Gruyter.
- Zgusta, L.** 1971. *Manual of Lexicography.* Janua Linguarum Series Maior 39. Prague: Academia / The Hague/Paris: Mouton.

Étude socio-terminologique du vocabulaire médical cilubà

Emmanuel Kambaja Musampa, *Département de Français–Langues africaines, Institut Supérieur Pédagogique, Mbuji-Mayi, République Démocratique du Congo (emmkamus@yahoo.fr)*

Résumé: À l'issue d'une enquête (menée à Mbuji-Mayi, en République Démocratique du Congo) auprès du personnel médical et paramédical locuteur et utilisateur du cilubà, langue bantoue, classifiée par Guthrie dans la zone linguistique L31, nous avons récolté, grâce au recours préalable au dépouillement documentaire, 150 termes médicaux français traduits en cilubà. Ces termes ont été soumis à une analyse socio-terminologique double: une analyse morphosyntaxique et une analyse sémantique. Il s'agit d'analyser une terminologie médicale, encore essentiellement véhiculée en français. Cette analyse aboutit à une structure morphosyntaxique et sémantique révélant que le vocabulaire médical en cilubà est «banalisé», il contient plus de termes simples de la classe 7 et ceux de la classe 12 que les autres. Ces termes sont polysémiques, métaphoriques, métonymiques, synonymiques, ou marqués par la synecdoque et le glissement sémantique. La monosémie n'est pas la règle, contrairement à la vieille tradition en terminologie. Cette étude s'insère dans la terminologie africaine et prépare ainsi à un dictionnaire médical bilingue français–cilubà.

Mots clés: TERMINOLOGIE, VOCABULAIRE BANALISÉ, STRUCTURE MORPHOSYNTAXIQUE, STRUCTURE SÉMANTIQUE, MÉTAPHORE, POLYSÉMIE, SYNONYMIE, MÉTONYMIE, MONOSÉMIE

Abstract: **Socioterminological Study of the Ciluba Medical Vocabulary.** At the conclusion of an investigation (led by Mbuji-Mayi, in the Democratic Republic of the Congo) among medical and paramedical staff who are speakers and users of Ciluba, a Bantu language classified by Guthrie in linguistic zone L31, we have collected, by means of recourse to the documentary evidence, 150 French medical terms translated into Ciluba. These terms have been submitted to a twofold socioterminological analysis: a morphosyntactical and a semantic analysis. It deals with analyzing a medical terminology, still basically conveyed in French. This analysis leads to a morphosyntactic and semantic structure which shows that the medical vocabulary in Ciluba has become general, containing more commonplace terms from classes 7 and 12 than from the other classes. These terms are polysemic, metaphoric, metonymic, synonymic or characterized by synecdoche and semantic shift. Contrary to the established tradition in terminology, monosemy is not the rule. This study deals with African terminology, in this way preparing for a bilingual medical dictionary French–Ciluba.

Keywords: TERMINOLOGY, GENERAL VOCABULARY, MORPHOSYNTACTIC STRUCTURE, SEMANTIC STRUCTURE, METAPHOR, POLYSEMY, SYNONYMY, METONOMY, MONOSEMY

1. Introduction

La présente étude a été menée dans le cadre d'un vaste projet sur la lexicographie des langues congolaises, et principalement la lexicographie du cilubà, langue bantou parlée au centre de la République Démocratique du Congo et classifiée par Guthrie dans la zone linguistique L31.

Cette étude, première de la série, est une analyse terminologique, vue sous un angle sociolinguistique. Il n'est pas ici question de créer des lexèmes pouvant être utilisés dans le domaine médical; encore que pareille entreprise n'est pas de moindre importance pour une langue non encore dotée de terminologie, dans presque tous les domaines (sauf le domaine religieux)! Seulement, pour l'heure, il nous a semblé important de décrire le vocabulaire médical créé par la communauté linguistique dont nous étudions le parler.

C'est pourquoi nous avons choisi de parler d'étude socio-terminologique. Ceci paraîtrait aux yeux de certains lecteurs un pléonasme vicieux dans la mesure où la terminologie se définit comme «la discipline qui étudie les vocabulaires spécialisés et en analyse les conditions d'utilisation» (http://www.cfwb.be/franca/pg_011.ht). À notre point de vue, il n'est nullement un pléonasme: nous étudions non seulement *les conditions d'utilisation* mais aussi *les mécanismes de création* du lexique spécialisé par la communauté.

En effet, il faut standardiser les termes déjà créés par la société, pour permettre la communication et l'assimilation, au lieu de créer d'autres concurrents. La langue est un phénomène social.

De ce qui précède, notre objectif est de décrire sur le plan à la fois morphosyntaxique et sémantique ce vocabulaire «banalisé»¹, vocabulaire fort utile parce que servant de trait d'union entre les médecins, le personnel médical ou paramédical et les non-spécialistes (patients ou autres).

Cette description jette les bases d'un lexique médical bilingue français-cilubà, que nous appelons de tous nos vœux. En outre elle aidera, tant soit peu, le traducteur bilingue (français-cilubà). Nous appliquons la méthode descriptive structurale à l'étude des données lexicales (Bal 1966: 124).

Ce travail n'est certes pas isolé, il s'inspire des lexiques déjà existants, publiés par le Centre d'études de linguistique théorique et appliquée de Kinshasa. Il s'en écarte à deux points de vue: du point de vue de l'objectif et du point de vue de la procédure.

Du point de vue de l'objectif, notre travail n'est pas un travail de vulgarisation du lexique médical, cherchant donc à répondre aux besoins des utilisateurs. Il est un effort de compréhension, par le terminologue-lexicologue de l'objet de son travail. Nous proposerons, à la fin, des options de création terminologique.

Du point de vue de la procédure, nous ne créons pas de termes, nous n'évaluons pas les équivalences français-cilubà, nous analysons les structures lexicologique et sémantique.

Notre méthodologie est la suivante: nous avons procédé à une enquête qui consistait à dire en cilubà 150 termes français recouvrant quatre domaines de la médecine, à savoir les pathologies, les actes médicaux, les lieux et les instruments médicaux. Ces lexèmes ont été tirés des ouvrages spécialisés.² Lors de ce dépouillement documentaire, nous avons privilégié les termes les plus courants.

Cette enquête a été menée auprès des locuteurs du cilubà prestant dans le domaine médical: deux infirmiers ayant le grade académique de gradué en Sciences et Techniques médicales, un infirmier technicien A2 (ayant le niveau équivalent à celui d'un D6), un médecin enseignant à l'Université de Mbujimayi.

Comme nous l'avons déjà dit ci-haut, nous utiliserons la méthode descriptive pour analyser nos données. Il s'agit d'une étude qui s'efforce de dégager les structures propres au lexique étudié, sur le plan morphosyntaxique et sur le plan lexico-sémantique.

À l'issue de cette brève introduction, nous proposons au lecteur la discussion des résultats. Il nous a semblé pratique de présenter en annexe les données.

Au cours de la présente étude, nous utiliserons les abréviations et représentations ci-après:

SN = syntagme nominal
SV = syntagme verbal
N = nom
SCon = syntagme connectif
SLoc = syntagme locatif
Loc = locatif
PaNA = participe nominal actif
PaV = participe verbal
V_{inf} = verbe à l'infinitif
Dét = déterminant
SP = syntagme prépositionnel
PaNP = participe nominal passif
Adj = adjectif
Adv = adverbe
Con = connectif
P = énoncé (phrase)
SAdjoint = syntagme adjectif
Inf = infinitif

Notez que, dans plusieurs contextes, la voyelle finale du mot lubà devient basse lorsqu'elle est suivie d'une voyelle basse. Dans le cas contraire, elle reste haute. Ceci est une loi phonologique et n'a aucune incidence sur le sens (Exemple: *kànsêra aku* (ce cancer là) et *kansêrà kàdi kàmùtâcisha* (le cancer le malmène)).

Au regard des résultats d'enquête, notre discussion reposera sur deux questions principales (chacune étant double):

- Quelle est la structure morphosyntaxique de ce vocabulaire? Quelle application terminologique on peut en faire?
- Quelle est la structure sémantique de ce vocabulaire médical? Quelle implication pour le travail terminologique?

Toutes choses étant égales par ailleurs, notre exposé s'étale sur deux axes:

2. L'analyse morphosyntaxique

Le lexique médical en cilubà est largement dominé par les termes de la classe 7. En effet, plus ou moins 21,05% des termes simples du corpus appartiennent à cette classe qui renvoie aux substantifs désignant les maladies (*cilundù*, *cyuwùjà*, *cyâdi*, *ciseki*, *cikupi*, *cifù*), l'état (*cyoyoyi*, *cyonà*), les objets (*cipimu*, *cileledi*, *cipandidi*) ou les résultats d'actions (*cisàlu*, *cipimu*)

De ce qui précède, lors de la création des termes, le terminologue doit le plus rechercher dans cette classe, car il y a plus de chances qu'il trouve «la perle recherchée» et que celle-ci vienne à s'adapter bien au système linguistique du cilubà. En procédant ainsi, non seulement il se conforme au génie de la langue, mais aussi il crée des mots que la communauté assimilera vite.

En outre, la classe 12, la classe du diminutif est aussi plus employée que les autres dans la terminologie médicale lubà (15,79%). Ceci est plus logique: il s'agit de l'idée de comparaison qui se situe derrière tous ces nouveaux termes. En comparant, par exemple, la particule d'appendice à l'intestin, il est clair que l'on désignera l'appendice par un diminutif en tant que prolongement du caecum.

Certaines classes n'apparaissent pas dans notre corpus: la classe 2 (le pluriel de la classe 1) et la classe de personnes. Cependant ce qui justifie cette absence est le fait que la personne peut aussi être exprimée par la classe 4. Ainsi, l'agglutination des préfixes des classes est compréhensible dans la mesure où les genres 1/2 + 4 désignent des personnes: on entend également «*bamingàngà* (les médecins/les infirmiers): le pluriel».

Il est alors très important pour le terminologue de marquer le genre et les autres modifications morphologiques à l'entrée de son article.

Quatorze lexèmes de classe 5 sont des gérondifs, il n'y a dans cette classe qu'un seul substantif (*dici*). Ceci est conforme au système du cilubà, dans lequel la plupart des mots de la classe 5 sont des gérondifs (formes nominales du verbe). Nous constatons que le cilubà préfère pour les actes médicaux, les états et les actions, recourir à l'infinitif.

Dans la terminologie médicale lubà, nous trouvons également les emprunts. La plupart d'entre ces néologies sont des calques (il y a aussi des adaptations: *bilààdi* et *lùpitaadi*). Tous ces emprunts ont été faits au français, la langue du personnel médical et para-médical.

Le terminologue ne peut s'empêcher de recourir à l'emprunt. D'ailleurs, en le disant ce n'est qu'une pétition des principes. Toutefois il ne doit pas se fier au transcodage du non-spécialiste!

Quant à l'accord de ces emprunts, il se fait normalement selon que le terme est classé dans l'une ou l'autre classe: *àmìbà*, *àlùkòlà*, *àpàdìsità*, *òpítalà*, *tròmpà*, *kìnìnà*, *làbo*, *kìlìnìkà* ayant la classe \emptyset prennent les préfixes *u-* et *i-*, c'est-à-dire en 1/4. Seuls les termes dont la première syllabe est interprétée comme préfixe de classe s'accordent selon lesdits préfixes: *Bilààdì bìdì bintàcìsha* (Le bilharziose me tourmente), *Dìspànsêrà dídì mutancì mule* (Le dispensaire est situé loin d'ici), *Kànsêra kàdì kàshebeya bantu bavulè* (Le cancer tue beaucoup de gens).

Il n'est pas sans importance de constater que ce vocabulaire emploie une variété de syntagmes nominaux (SN):

SN1 = N + SN2
Con + SN1
N + SCon
N + SLoc
N + Adj
N + PaNA
N + N
Con + N + PaV

Au regard de cette variété, le terminologue dispose d'un éventail de choix pour la création des syntagmes. Les structures que nous avons découvertes sont plus variées pour permettre la création syntagmatique.

Les syntagmes verbaux (SV) ne sont pas eux non plus absents du vocabulaire médical lubà. Dans notre corpus, ils sont aussi en bon nombre (22). Toutefois, par rapport au SV, le SV n'est pas aussi varié. Il n'existe que deux types de SV: SV à noyau gérondif et SV à noyau infinitif.

Le noyau verbal (gérondif ou infinitif) est suivi d'un complément qui peut être un N, un SN, un SV, un SAdjoint (le circonstanciel).

Notre corpus a aussi révélé le recours à des énoncés pour exprimer les référents médicaux. Certes, ces énoncés ne sont pas nombreux, mais ils présentent une variété au niveau de la base prédicative: tantôt on retrouve une base prédicative, tantôt une base participe nominal, tantôt un participe verbal.

Pour clore cette description morphosyntaxique, nous disons que le vocabulaire médical ici analysé présente plus de formes nominales que de formes verbales et des énoncés. Conséquemment, le terminologue devra privilégier le recours aux formes nominales (principalement les substantifs simples et les syntagmes nominaux).

3. L'analyse sémantique

3.1 La métasémie

Nous examinerons quelques traits métasémiques de notre vocabulaire: la métaphore, la polysémie, la synecdoque, la métonymie, la synonymie, le glissement morpho-sémantique.

— La métaphore

La métaphore est l'une des réalités fondamentales du langage humain. Aucun lexique ne peut être sans métaphore. Car elle n'est pas le fait de la rhétorique, elle est plutôt présente dans toute notre parole. Les terminologues, bilingues ou monolingues, en font quotidiennement l'expérience. À titre de rappel, disons avec Kleiber (1983, cité par Meyer 1988) qu' :

il n'y a métaphore que si: (1) le sens littéral ne correspond pas à ce que qu'a voulu dire le locuteur/ou ne renvoie pas à un référent qui fait partie de sa référence virtuelle, (2) la compréhension de ce qu'a voulu dire le locuteur/ou la découverte du référent actuel «étranger» passe par la mise en application de mécanismes de similitude (<http://www.info-metaphore.com/definition.html>).

Ajoutons avec Dubois (1975: 202) que la métaphore, considérée comme intersection de deux ensembles sémiqiques, comporte trois unités sémantiques (signifié de départ «D», signifié d'arrivée «A», signifié intersectif «I») dont aucune n'est, en théorie, prééminente par rapport aux autres.

Les termes suivants sont métaphoriques: *diboba*, *kalòwa*, *kashèlèlèkà*, *buzevu*, *dyuminyina*, *dìsaamà dyà diboba*, *dibùùkà dyà mashi*, *dijìmina dyà lungènyi*, *mukàjì kuulu kula*, *kampala dibàkà*, *kubàndisha mubidi*, *kusabula mwâna*, *kutùùla difù*.

Nous avons rangé ces métaphores en quatre catégories:

- les métaphores du devenir,
- les métaphores de la mesure,
- les métaphores de l'espace, et
- la métaphore de la forme.

Notre classification repose sur le principe d'existence d'une catégorie fondamentale, prototypique, autour de laquelle tournent toutes nos métaphores en tant que systèmes de pensée. Nous soutenons le point de vue de Lakoff et Johnson, selon lequel la métaphore est inscrite dans la pensée: «Notre système conceptuel ordinaire, qui nous sert à penser et à agir, est fondamentalement métaphorique» disent-ils (Lakoff et Johnson 1980: 13).

— Les métaphores du devenir

Sous cette catégorie, nous rangeons les métaphores suivantes: (1) *diboba* (kwashiorkor); (2) *dyuminyina* (convulsion).

Le changement d'état est une réalité vitale. Pour l'exprimer, le vocabulaire médical, que nous analysons, recourt à la métaphore du devenir. L'énoncé (1) renvoie à un passage du cru, du fondamentalement premier, de la couleur première à ce qui a perdu son premier état sous l'effet d'un quelconque agent. Une personne en bonne santé, un objet en bon état sont comparés à un fruit cru. L'on entend dans le langage courant: *màshinyi mabishi*, *mutèlù mubishi*, *nsongà-kàjì ùcìdì mubishi*. Littéralement on traduira: véhicule cru (pour dire un véhicule neuf), une chemise crue (pour dire une chemise neuve), une fille encore crue

(pour dire une fille vierge).

Diboba se dit pour une personne qui présente les signes évidents de malnutrition. La pensée littérale est celle-ci: le corps de ce malade est passé (de vient) de sa couleur première, la couleur de la vitalité, de la rudesse comme le cru, à la couleur d'un objet cuit, sous l'effet de la chaleur (ou du feu).

Une autre métaphore du devenir est la (2). Il est ici question du passage de ce qui était pliable à ce qui est devenu non-flexible, dur au toucher et allongé, sans vie, donc. L'énoncé *mwâna ñmmuuminyina* se traduit littéralement par *l'enfant est devenu non-flexible. Il ressemble ainsi à du bois sec, un bois sans vie.

Comme on le voit, les deux métaphores du devenir s'inspirent des réalités d'autres domaines pour exprimer celles qui sont du domaine médical. Le passage du cru au mûr relève du domaine des objets (surtout d'aliments) et le passage du flexible (de l'humide ou du mou) vers le non-flexible (le dur) relève du domaine de la matière.

— Les métaphores de la mesure

Buzevu est l'une des illustrations de la métaphore in absentia attestée dans le vocabulaire médical sous analyse. Elle est, heureusement, une métaphore de la mesure. En effet, ce terme renvoie à l'éléphantiasis qui est augmentation considérable du volume d'un membre ou d'une partie du corps due à l'obstruction des vaisseaux lymphatiques. En régions intertropicales, la cause majeure d'éléphantiasis est la filariose lymphatique, transmise par les moustiques (voir <http://www.chu-rouen.fr/ssf/pathol/elephantiasis.html>).

Dans le parler lubà, ce gonflement est comparé à la taille de l'éléphant (*nzevu*). Morphosémantiquement, le terme pourrait être analysé en *bu-* (le morphème de classe 14) et *-zevu* (le thème nominal). C'est une heureuse équivalence entre la désignation et le référent, étant donné qu'il y a ici une motivation linguistique.

Le terme *kashèlèlèkà* renvoie d'abord à une pathologie qui se manifeste par le blocage de la croissance de l'enfant, accompagné des amaigrissements, la perte d'appétit, les pleurs sans arrêt. Cette pathologie est souvent associée à un mauvais sort. Jusqu'à ce jour, elle se traite mieux chez des femmes supposées détentrices du pouvoir de jeter et de lever ce sort. La pauvre victime de ce maléfice maigrit au point de rester la peau collée aux os, les veines amplement visibles, on dirait des filaments.

Ce terme renvoie à la réduction de la mesure, mieux à la réduction de la taille. On trouve dans le langage courant un autre adjectif du même radical — *shèè* — *kashèpètèka/mushèpètèka* (pour dire devenu de très courte taille; lorsqu'on parle d'une personne, d'un animal). Ainsi donc, nous comprenons pourquoi ce terme est aussi utilisé pour désigner le sida. Un adulte ou un enfant atteint du sida (*kashèlèlèkà*) ne cesse de maigrir.

— Les métaphores de l'espace

Le vocabulaire médical exprime certains référents au moyen des métaphores de l'espace ci — après (retrouvées dans notre corpus): (1) *kuulu kula* (la grossesse:

littéralement «être suspendu plus haut»), (2) *kusabula mwâna* (faire accoucher par césarienne: littéralement «sortir d'un trou plein d'eau un enfant»), (3) *dibù-ùkà dyà mashi* (l'anémie: littéralement «l'envol du sang»), (4) *kubàndisha mubidi* (faire la fièvre: littéralement «hausser le corps»).

La métaphore (1) renvoie à l'espace — hauteur. La femme qui est grosse est comparée à une personne située à une très haute altitude. Ceci suffit pour comprendre toute la peur qui entoure l'accouchement. (Disons au passage qu'on entend toujours dire les femmes: *kudilela sè ìku lufù*; littéralement traduit par: qui va à l'accouchement s'expose à la mort. L'accouchement est un danger capable de donner la mort. La femme qui attend famille se sent toujours en «insécurité», car le haut est insécurisant, tandis que le bas est rassurant. Lorsque la femme s'accouche, on entend dire: *mukàjì wàátùùluki bilenga*; littéralement: la femme est bien descendue. La métaphore de la hauteur est très forte: le haut n'est pas rassurant, le bas l'est au contraire. Le haut est synonyme de danger de mort permanent.

La métaphore (2) renvoie à l'espace — trou. Son équivalent français césarienne est une métonymie. *Kusabula mwâna* désigne littéralement: retirer l'enfant d'un trou dans lequel on trouve également de l'eau. Cette métaphore renvoie à l'idée que l'enfant baigne dans une cavité (ici l'utérus) dans laquelle il y a de l'eau (le liquide amniotique). Un autre sème attaché au terme *kusabula mwâna* est celui de «danger». L'enfant que l'on tire «par césarienne» est supposé avoir été en danger de mort, car ne pouvant être expulsé normalement.

La métaphore (3) renvoie à l'espace — étendu, à travers lequel l'intelligence, ici considérée comme un objet, est égarée. *Kujìmija lungènyi* (littéralement: égarer l'intelligence) s'oppose à *kupeta lungènyi* (re)trouver l'intelligence, littéralement: pour dire retrouver l'exercice de sa faculté mentale). L'on considère que celui qui développe des troubles «égare ses facultés mentales» ou même que celles-ci sortent de lui. D'où une métaphore que l'on entend couramment *kupingaja lungènyi* (littéralement: retourner l'intelligence, à sa place).

La métaphore (4) est aussi une métaphore de l'espace — hauteur, comme la (1). *Dibùùkà dyà mashi* est traduit littéralement par: l'évaporation du sang. Le sang s'évapore vers le haut. Là où le français exprime la privation (an-), le cilubà exprime la fuite, la perte. Le terme *dibùùkà dyà mashi* désigne l'anémie, due à une fièvre (souvent). Lorsque l'anémie est due à une autre cause, on parle de *dijika dyà mashi* (littéralement traduit par: l'épuisement du sang). Il n'est pas faux de dire que le sang est considéré comme l'eau (par son trait «liquide») chauffée par un pathogène à une température telle que l'évaporation est possible. C'est pourquoi, la fièvre est désignée par la température (*mubidi kapyà, mubidi mupyà*); littéralement l'on traduirait par: le corps est chaud.

— La polysémie

Le terme «polysémie» se définit toujours par opposition à la monosémie et à l'homonymie. Nous ne voudrions pas, pour l'heure en tout cas, nous replonger dans le vieux débat sur ces aspects des problèmes sémantiques. Nous pensons

que le lecteur intéressé, pourrait lire les ouvrages de Kleiber (1999) et Pergnier (1993) cités à la bibliographie de la présente étude.

Somme toute, nous définissons la polysémie comme le fait qu'un mot revêt une pluralité des sens, lesquels sont unis par un certain rapport. Pareille définition nous permet de nous éloigner de l'homonymie et des prototypes. Elle nous permet également d'éliminer de notre champ d'étude, les mots vagues.

Avant d'entrer dans le vif de l'analyse, disons qu'il est un fantasme dont nous devons nous défaire: celui de la monosémie de la terminologie scientifique. Notre corpus prouve (je pense que d'autres, même les plus hautement construits, le prouveront encore) que les termes peuvent aussi être polysémiques.

La polysémie tient de la nature même du signe linguistique. Le terminologue doit cesser de ramener la signification au sens et savoir que le terme peut désigner plusieurs référents selon les «contextes».

Venons-en aux polysèmes du corpus: (1) *mâyi*, (2) *diteeta/mateeta*, (3) *musàndà*, (4) *cipimu*, (5) *difû*, (6) *mungàngà*, (7) *mubidi*.

— *mâyi*: Seul le contexte en précise le sens.

Dans l'énoncé (1) *Bàdi bàmwela mâyi bwalu ùdi nè màlàriya* (Il est sous perfusion parce qu'il fait de la malaria), *mâyi* renvoie au soluté de perfusion.

Dans un autre énoncé du genre (2) *Mungàngà ùdi ùkòka mâyi àà mu mwongu* (L'infirmier fait une ponction lombaire), le terme *mâyi* désigne le liquide céphalo-rachidien.

Dans un troisième énoncé (3) *Mwâna ewu ìmmunwà mâyi pa kulediibwa* (Le bébé a inhalé le liquide amniotique à la naissance), le terme *mâyi* désigne le liquide amniotique.

Il signifie le sperme dans le contexte suivant: (4) *Bâyendà kèna nè bilùmà mu mâyi èndà to* (Le sperme de son mari manque des spermatozoïdes, c'est-à-dire ce qui s'appelle azoospermie).

On entend lors des séances d'éducation sanitaire, au sujet du traitement de la diarrhée, l'infirmier dire (5) *Pàdi mwâna wèla munda, bìdi bìkèngela kumunwìsha mâyi misangu mivulè* (Quand un enfant fait de la diarrhée, il faut lui donner régulièrement de l'eau à boire, c'est-à-dire pour prévenir la déshydratation).

Enfin, le terme *mâyi* dans le dernier contexte, (6) *ùdi nè mâyi ku bisùlusulu* (Il a du liquide dans la cavité pleurale), désigne le liquide qui emplit la cavité pleurale.

Comme on le voit, le terme polysémique *mâyi* ne peut nullement être confondu: le contexte précise toujours le sens. Patients et personnel de santé communiquent sans aucun heurt.

— *diteeta/mateeta*: Ce terme a plus d'une désignation. Dans l'énoncé (1) *Ndi mufila bunyau bwà mungàngà ènzà mateeta* (J'ai donné les selles au laboratoire pour examens), il désigne les analyses des échantillons au laboratoire.

Dans un autre énoncé du type: (2) *Ngaayi nè mwâna kwà mungàngà, wèénjì mateeta mubidi mujimà, kèna musangàna cintu* (Le médecin a ausculté tout le

corps de l'enfant que je lui ai amené, mais hélas, il n'a rien trouvé!), le terme désigne l'auscultation. Dans un cas comme dans l'autre, le terme désigne le fait d'examiner un échantillon ou un malade.

— *musàndà*: Ce terme est également polysémique. Nous pouvons le voir à travers ces énoncés:

(1) *Mpùnga udi nè misàndà; ke bwalu kaayi ùdi nè difù dipàta* (Mpùnga a des verminoses, raison pour laquelle il présente un ballonnement de ventre).

(2) *Mpùnga ùdi ùnàya ku makàsà kutupù mu mvùla, neàdyatà misàndà* (Par le fait qu'il joue pieds nus sous la pluie, Mpùnga aura à marcher sur des vers de terre).

Dans les deux énoncés, le terme *misàndà* (plur. de *musàndà*) n'a pas le même sens: il signifie tantôt «verminose» tantôt «vers de terre». C'est pareil aussi pour le français.

— *cipimu*: Ce terme désigne tantôt la pesée, tantôt la consultation prénatale, tantôt le prélèvement de la température. C'est cela que nous comprenons à travers les énoncés suivants:

(1) *Ndààku nè mwâna ku cipimu* (Amène l'enfant à la pesée).

(2) *Pàdi mukàji nè difù, bidi bikèngela ìkalà ùya ku cipimu* (Il est recommandé à la femme qui attend famille d'aller régulièrement à la consultation prénatale).

(3) *Cipimu cyà mwâna còwa àmu mu makùmi ànaayi* (La température de l'enfant était stationnée à 40°C).

— *difù*: Il désigne tantôt la grossesse, tantôt le ventre, tantôt les coliques. Soient les énoncés:

(1) *Kapinga ùdi nè difù* (Kapinga est grosse).

(2) *Mungàngà ùdi ùbaabata mu difù dyà mubèèdi* (Le médecin tâte le ventre du malade).

(3) *Difù didi òmunyènga* (Il a des coliques abdominales). Cet énoncé est l'équivalent de *munda mùdi òmunyènga* qui se traduit littéralement par: Ce qui est à l'intérieur du ventre me torsionne.

Comme on le constate à travers ces énoncés, le contexte discursif et phrasique lève toute possibilité éventuelle de confusion.

— *mungàngà*: Ce terme désigne tantôt docteur/médecin, tantôt l'infirmier, tantôt le laborantin, tantôt le tradi-praticien. Nous le voyons dans les énoncés ci-après:

(1) *Mungàngà wàátèèki mwâna panshi* (L'infirmier a circoncis l'enfant). *Mungàngà* a le sens d'infirmier.

(2) Dans *Mungàngà wààngàci mashi àà mu munu* (Le laborantin a fait la goutte épaisse), le terme désigne le laborantin. En parlant à un médecin, le terme a le sens de médecin ou de docteur: *Mungàngà wânyi, ndi nè cyâdi kùdi ngondu isambòmbò* (Cher docteur/médecin, voici six mois que j'ai la tuberculose). Le terme est aussi utilisé pour désigner le pharmacien. Enfin, *mungàngà* est employé pour quiconque soigne. C'est pourquoi il s'emploie pour le tradi-

praticien aussi (celui qui soigne à base de plantes comme dans la tradition; sans avoir été dans une école de médecine).

— Le terme *mubidi* désigne tantôt la fièvre, tantôt la santé, tantôt le corps lui-même.

Dans l'énoncé (1) *Mwâna ùdi ùsaama mubidi*, le terme désigne la fièvre. Cet énoncé se traduit par: L'enfant fait la fièvre. Mais admettons qu'on se trouve à une séance de vaccination, l'infirmier pourrait dire à une maman: (2) *Mwanèbà kèna nè mubidi mwîmpà to*. Dans ce contexte, le terme désigne l'état de santé. Cet énoncé se traduirait par: Ton enfant n'a pas bonne santé.

L'énoncé (3) *Mbîmpà kulama mubidi wèbà nè mankendà bwà kwepuka twîshi*, dit par une infirmière à une maman lors de la consultation prénatale, contient le terme *mubidi* qui désigne le corps. Il se traduirait par: Il est mieux de garder ton corps propre en vue d'éviter des infections (en cette période de grossesse).

Enfin, par euphémisme, le terme *mubidi* est employé pour désigner le sexe ou les rapports sexuels: (4) *Mwanèbà wa balùma eu, ùkaadi mumanyà mubidi wà muntu mukàjì pàdîye apa* (Ton fils que voici a déjà eu des rapports sexuels avec une femme).

Comme on peut le constater, la polysémie lexicale est un fait attesté dans le lexique médical. Une note cependant est à faire: la polysémie n'affecte pas les syntagmes nominaux (ou verbaux). Elle est limitée aux mots dits simples: la longueur de l'unité lexicale lève l'ambiguïté et la polysémie.

Dans le même ordre d'idées, nous analyserons également le phénomène de synecdoque.

— La synecdoque et la métonymie

Nous étudions les deux métasémies ensemble parce que la première n'est qu'un cas particulier de la deuxième. Pour rappel, la métonymie consiste à nommer un concept au moyen d'un terme désignant un autre concept, lequel entretient avec le premier une relation d'équivalence ou de contiguïté. Et la synecdoque est une métonymie particulière qui consiste à considérer tout pour la partie, la partie pour le tout, un élément pour l'ensemble.

En effet, synecdoque et métonymie ne sont pas à rechercher dans le vocabulaire médical lubà. Les termes *mucîma*, *dicî*, *cifû*, *cyâdi*, *munda*, et *mutû* sont des synecdoques. *Difû* et *mubidi* sont aussi métonymiques.

Certaines maladies sont exprimées en termes de parties où elles se localisent. Ainsi trouve-t-on *mutû* (littéralement: tête) pour désigner les céphalées ou les migraines, *mucîma* (littéralement: le foie) pour désigner l'hépatite A, B, C ou la cirrhose de foie, *dicî* (littéralement: oreille) pour désigner toutes les maladies de l'oreille, *cifû* (littéralement: estomac) pour désigner la gastrite, *cyâdi* (littéralement: la poitrine) pour désigner la tuberculose, *munda* (littéralement: l'intérieur) pour désigner la diarrhée ou autres pathologies internes. Les termes *difû* et *mubidi* sont également métonymiques. *Difû* qui se traduit littéralement par ventre peut, dans un autre contexte, comme le suivant, désigner la grossesse:

(1) *Kapinga ùdi nè difù* (Kapinga est grosse). C'est la métonymie du contenant pour le contenu. En effet, toute personne, mâle ou femelle, a un ventre (*difù*). Cependant, dans le cas d'espèce, il s'agit de l'enfant logé dans l'utérus, lequel est placé dans le petit bassin (ou pelvis) féminin. Au cours de la croissance de l'embryon, puis du fœtus, le volume du ventre augmente. Alors, le locuteur du cilubà parle du ventre (le contenant) au lieu du fœtus, le contenu.

Enfin le terme *mubidi* est aussi à la fois métaphorique et polysémique. Dans l'énoncé *Mwâna ùdi ùsaama mubidi* (L'enfant fait de la fièvre), le terme *mubidi* désigne la température du corps. Il est dans ce cas métonymique.

— La synonymie

C'est bien étrange de parler de synonymie dans un vocabulaire médical (une terminologie). Cependant, il ne serait pas honnête de notre part de ne pas en parler, car elle existe. Au cours de notre analyse du corpus, un seul cas de la synonymie s'est révélé.

Disons de prime abord que la synonymie se classe selon quatre types (d'après Kerbrat-Orecchioni 1979: 168-169):

- la synonymie complète et totale (extrêmement rare),
- la synonymie complète non-totale, e.g. beau-fils/gendre,
- la synonymie totale non-complète (extrêmement fréquente, la connotation qui oppose les deux synonymes pouvant être de nature extrêmement diverse), e.g. jaunisse/ictère (terme commun/terme technique); soixante/+ septante (connotation géographique); cacher/celer (connotation archaïque), et
- la synonymie non-totale ni complète: la plus fréquente, e.g. fermer/clore.

Il ressort de cette classification que le seul cas de synonymie attestée dans notre corpus est double. D'abord, il est celui d'une synonymie complète et totale: elle est attestée pour le triplé *mukupa/dibùluka/dipaala*. C'est cette synonymie que définit Ullmann quand il dit: «on ne peut appeler synonymie que les mots qui peuvent se substituer l'un à l'autre dans un contexte sans le moindre changement dans la portée cognitive ou affective» (cité par Kerbrat-Orecchioni 1979: 167).

On peut dire, en interchangeant les trois premiers termes que nous avons le même énoncé:

- (1) *Ilunga ùdi nè dipaala, ke bwalu kaayì ùdi ùdya byà ku diyaala.*
- (2) *Ilunga ùdi nè dibùluka, ke bwalu kaayì ùdi ùdya byà ku diyaala.*
- (3) *Ilunga ùdi nè mukupa, ke bwalu kaayì ùdi ùdya byà ku diyaala.*

Ces trois énoncés se traduisent par un seul «équivalent»: *Ilunga est devenu fou, c'est pourquoi il mange aux ordures.*

Ensuite, *katombùtombù*, le dernier terme est, par rapport aux trois précédents, en synonymie totale mais non-complète. Car la folie dont il est question

pour *katombùtombù* est une folie intermittente. Ce qui explique qu'il se trouve des contextes dans lesquels ce terme ne peut être interchangeable avec les trois autres.

— Le glissement morpho-sémantique

Nous désignons ainsi le sens qui se manifeste lors du passage du singulier vers le pluriel dans le cas des termes: *kîshi*, *twîshi*, *dîsaamà dyà mashi* et *mâsaamà àà mashi*.

Les termes *kîshi* et *twîshi* ne sont pas identiques sémantiquement, dans certains contextes. Nous en voudrions pour illustrations les énoncés:

- (1) *Bàyeebà ùdi nè kîshi* (Ton mari est sidéen/séropositif/tuberculeux).
- (2) *Bàyeebà ùdi nè twîshi mu mênyi* (Ton mari a des microbes dans les urines).
- (3) *Mbangùla kîshi ku mênyi* (On a découvert, à l'issue de l'examen d'urine, que la personne était séropositive).
- (4) *Mbangùla twîshi ku mênyi* (Les microbes ont été découverts après examen d'urine).

Dans le premier énoncé, le terme *kîshi* peut désigner le sida, ou la tuberculose (les deux peuvent apparaître en association). Ce qui ici le sème commun est le caractère «très contagieux». Donc, dans ce contexte, *kîshi* n'est pas le singulier de *twîshi*. Il ne peut accepter, d'ailleurs, le pluriel. Cet énoncé se traduirait par: Ton mari est séropositif (ou tuberculeux).

Dans le deuxième énoncé, *twîshi* désigne les germes pathogènes en général, les microbes. L'énoncé (2), en cas d'une infection urinaire, se traduirait par «ton mari a des infections urinaires». De ce qui précède, (3) se traduirait: «on a découvert le sida lors de l'analyse des urines» et (4) «on a découvert les microbes responsables d'infections urinaires lors des analyses».

Dans le même ordre d'idées, *dîsaamà dyà mashi* ne peut être considéré, dans tous les cas comme le singulier de *mâsaamà àà mashi*. Prenons les énoncés ci-après:

- (5) *Mwâna ìmmufwà dîsaamà dyà mashi*.
- (6) *Mâsaamà àà mashi àdi àshipa bâna bikolà*.

L'énoncé (5) se traduirait par: «L'enfant est décédé d'une drépanocytose» alors que le (6) se traduirait par: «Les maladies liées au sang tuent beaucoup d'enfants». Si dans l'énoncé (5) le singulier est précis, dans l'énoncé (6), le pluriel renvoie au général, de la leucémie à la drépanocytose.

3.2 La monosémie

Nous avons choisi de parler de la monosémie en passant parce que, du point de vue sociolinguistique, elle est une exception qui n'infirme pas la règle selon laquelle les lexèmes sont principalement polysémiques.

Certes, dans notre lexique, il n'a pas manqué des lexèmes monosémiques. Ils sont aussi nombreux (tous ceux qui n'ont pas fait l'objet de la métasémie). Il nous semble que la raison de la monosémie est la recherche de la rigueur dans la terminologie, caractéristique essentielle pour toute discipline scientifique.

3.3 Propositions de traduction terminologique

À l'issue de l'analyse ci-dessus, il nous revient de proposer quelques traductions, sans aucune prétention à la perfection. Ces propositions sont suivies, chacune de petites explications socioterminologiques. Mais il n'est pas ici temps de nous étaler sur les différentes techniques traductologiques utilisées. Nous avons retenu 18 termes:

(1) **méningite:** *cisèbasebabongù* 7/8

Ce terme est une lexie composée qui peut s'analyser en *cisèba* qui désigne le terme méninge. Ce terme est déjà connu dans la langue lubà dans laquelle il désigne peau. Dans ce contexte il renvoie à la membrane qui recouvre le cerveau, désigné par le terme courant *bongù*. Il y a ici une duplication du terme *cisèba* pour signifier l'inflammation de la membrane (littéralement: la membrane du cerveau qui grossit). Nous avons mis ensemble les lexèmes: *cisèbaseba cyà bongù*. Le connectif a été supprimé.

Dans le même ordre d'idées les termes de la même famille se diraient, pour le méningisme: *busèbasebabongù* 14/0; méningiose se traduirait par *seba-bongùnènè* 0/4 qui signifie en général toute inflammation de méninges.

(2) **pédiatrie:** *bungàngâna* 14/0 qui s'analyserait en *bungàngà* (la médecine) *bwà bâna* (des enfants) avec le thème nominal *-âna* (signifiant enfant). De la sorte le terme chirurgie se traduirait par *bungàngàpandi* 14/0 (la médecine d'opération). La gynécologie serait *bungàngànkâji* 14/0 (la médecine des femmes) et la médecine interne se traduirait par *bungàngàmùndà* 14/0 (la médecine de l'intérieur).

(3) **globule:** *kabùlùngù* 12/13. Ce terme existe déjà en cilubà et désigne un petit objet rond. Il nous semble approprié pour désigner «globule». Dans cet ordre d'idées, le terme «globule rouge» sera désigné par *kabùlùngù mashi*. Ce composé désigne directement le sang car le globule rouge est la cellule de sang qui contient l'hémoglobine désigné par métonymie, en cilubà, par le terme *mashi* (le sang).

(4) **hyperglycémie:** *bòdyòdyòsùkâdì* 0/4. En effet, en cilubà, le verbe *kubòdyodyoka* désigne proliférer, être en grande quantité. C'est bien là l'idée d'une hyperglycémie, terme qui renvoie à l'excès de glucose dans l'organisme. Il est aussi un composé réunissant *bòdyodyoka* (le verbe) et le substantif *sùkâdì* (sucre).

(5) **hypoglycémie:** De ce qui précède, il y a lieu de créer le terme *shèèpèsùkâdì* 0/4 composé de la racine *-shèèpe-* qui désigne le fait de se réduire, de diminuer de taille, de devenir petit. Ce qui arrive effectivement dans le cas d'une hypoglycémie.

(6) **hypocalcémie:** *shèèpèkàlisyumù* 0/4 dans lequel nous retrouvons la racine *-shèèpe-* dont nous venons de parler et *kàlisyumù* 12/13 est le terme courant utilisé par les locuteurs du cilubà, terme emprunté au français pour désigner le calcium. De la sorte la calcémie se traduirait par *bùlisyumù* 14/4.

(7) **leucémie:** se traduirait par *mbùdikabùlungù* 1/4, lexème composé à partir *kubùdika* (se multiplier, devenir en surnombre) et *bùlungù de ka/tu — bùlungù* (terme par lequel nous avons traduit «globule»). La leucémie est ainsi entendue comme une prolifération des globules blancs ou autres dans le sang.

(8) **virus:** *kìshincii* 12/13 lexème composé de *kìshi* (petit insecte) et l'adjectif *ncii* qui signifie «très petit». Dans cet ordre d'idées, le virus est défini comme un petit (très petit alors) organisme qui demeure dans la cellule.

(9) **hémoglobine:** *mpùlùnkùnzù* 1/4. Ce terme est un lexème composé de *mpùlù* qui désigne un petit fruit rouge dit en cilubà *mpùlùmànjà* et *kùnzù* qui est adjectif de couleur désignant la couleur rouge. Pour dire que l'hémoglobine est une protéine qui donne sa couleur rouge au sang.

(10) **anémie:** *shèèmashi* 0/4. Ce terme est lié à ceux que nous avons déjà créés plus haut sur base de la composition verbo-adjectivale *-shèè-*: diminuer, s'épuiser, et *mashi* qui désigne dans le langage courant «le sang».

(11) **antibiotique:** *mulùngwìshì* 3/4, un terme composé de deux substantifs, à savoir *mulùngù* (poison) et *-ìshi* (insecte dont nous avons déjà parlé (voir **virus**)).

4. Conclusion

À quelle conclusion nous a conduit cette étude socioterminologique?

Nous sommes arrivé à des résultats tels que le vocabulaire médical de la communauté linguistique lubà, dans son état actuel, présente la structure sémantique et même morpho-syntaxique de la langue courante. De ce fait, c'est un vocabulaire banalisé.

Sur le plan terminologique, il est conséquent que le terminologue se débarrasse des vieux fantasmes qui l'empêchaient de voir les métasémies qui jalonnent la terminologie. La métaphore, par exemple, est inscrite, non dans la langue, mais dans notre système de pensée.

Certains traits sont fréquents, d'autres non. Au terminologue de s'inspirer de cette répartition morpho-syntaxique et lexico-sémantique pour son travail sur la terminologie médicale en lubà élaborée systématiquement.

Nous avons nous-même proposé une terminologie constituée de 19 termes en vue de montrer les voies à suivre en vue d'un travail lexico-terminologique utile. Notre espoir est que plusieurs terminologues du domaine lubà s'engageront dans cette voie.

Notes

1. Par vocabulaire banalisé, nous entendons le vocabulaire accessible à une population plus ou moins nombreuse, mais nécessairement plus nombreuse que celle qui possède la compétence du langage spécialisé dont il est issu, et moins nombreuse que celle qui ne connaît que le langage courant (Galisson 1979: 125)
2. Il s'agit de Bernard et Pierre (1989), Janssens et Courtejoie (1996) et *Manuel du centre de santé* (1988).

Bibliographie

- Bal, W.** 1966. *Introduction aux études de linguistique romane — avec considération spéciale de la linguistique française*. Paris: Didier.
- Bernard, G. et G. Pierre.** 1989. *Dictionnaire médical pour les régions tropicales*. Kangu-Mayumbe: BERPS.
- CHU Hôpitaux de Rouen.** 18 février 2006. *Le Catalogage et l'Indexation des Sites Médicaux Francophones (CISMeF): éléphantiasis*. <http://www.chu-rouen.fr/ssf/pathol/elephantiasis.html> [17 mai 2006].
- Detienne, C.** 23 février 2006. *La métaphore en question*. <http://www.info-metaphore.com/definition.html> [17 mai 2006].
- Galisson, R.** 1979. *Lexicologie et enseignement des langues*. Paris: Hachette.
- Guthrie, M.** 1948. *The Classification of Bantu Languages*. Londres: Oxford University Press.
- Janssens, P. et J. Courtejoie.** 1996. *Aide-mémoire pour le dispensaire*. Kangu-Mayumbe: BERPS.
- Kerbrat-Orecchioni, C.** 1980. *L'énonciation: de la subjectivité dans le langage*. Paris: Armand Colin.
- Kleiber, G.** 1999. *Problèmes de sémantique — La polysémie en question*. Paris: Presses universitaires du Septentrion.
- Lakoff, G. et N. Johnson.** 1980. *Les métaphores dans la vie quotidiennes*. Paris: Éd. De Minuit.
- Manuel du centre de santé. Vol. I: Soins curatifs*. 1988. Kinshasa: Projet de santé pour tous.
- Pergnier, M.** 1993. *Les fondements socio-linguistiques de la traduction*. Édition remaniée. Lille: Presses universitaires de Lille.

ANNEXE 1: Classement des termes

1. Les formes nominales

1.1 Les substantifs

1.1.1 Selon les classes

Cl 1	Cl 2	Cl 3	Cl 4	Cl 5	Cl 6	Cl 7	Cl 8	Cl 11	Cl 12	Cl 13	Cl 14	Cl 15	Cl 16	Cl 17	Cl 18	Tot.
mu-	ba-	mu-	mi-	di-	ma-	ci-	bi-	lu-	ka-	tu-	bu-	ku-	pa-	ku-	mu-	
5	0	6	4	4	3	12	2	6	9	2	3	0	0	0	1	57

1.1.2 Selon le genre

1/0	1/4	0/4	5/6	5/0	0/6	7/8	0/8
mfwènka nkooyi mpalù nsàdi mpùta	mungàngà (ngàngàbukà) mutù munànà mukupa mucìma musàndà	milubàlubà mpùsù misònji mbàlàngà	diteeta dici	dibùluka dipaala	màsàsà mâyi mankùbùngu	cilundù cyoyoyi cyonà cipimu ciseki cikupi cifù cileledi cipandidi cyâdi cyuwùjà cisàlu	binyòka bilùma
5	6	4	2	2	3	12	2

11/4	11/0	12/13	0/13	14/0	14/6	18/0	Total
lusòngà lusundu lukosù lushìngì	lùbùngù londàpu	kalwalwalwa kishi kakèla katombùtombù kalòwa kàntembela kambulu kashèèlèlèkà kasèndà	twîshi (selon le contexte) tunèka	buzevu bupofu	bulààlu	munda	
4	2	9	2	2	1	1	57

1.2 Les gérondifs

Nous avons recensé 11 gérondifs appartenant au genre 5/0:

- diboba
- dikùluka
- dilènga
- dilèwuluka
- dipaluka
- dipanda
- ditàpika
- diteeta
- dyondapa
- dyuminyina
- dyumusha

1.3 Les emprunts

Il s'agit de termes que l'on reconnaît encore comme étrangers, car il faut reconnaître que beaucoup de mots étrangers entrés dans le lexique lubà fonctionnent tellement bien que même certains locuteurs natifs ont du mal à les reconnaître comme tels. Les 10 emprunts que nous avons retrouvés dans notre lexique sont:

- àlùkòlà
- àpàndisità
- bilààdi
- dìspan-sêrà
- kànsêrà / kànsêlà
- kilinikà
- làbo
- kìninà
- òpìtalà / lùpìtaadi
- tròm-pà
- àmìbà

Presque tous appartiennent au genre 1/4, à l'exception de bilààdi (8/8), dìspan-sêrà (5/6), kànsêrà (12/13)

1.4 Les syntagmes nominaux

1.4.1 Déterminé + déterminant

- dìsaamà dyà mwoyi / dìsaamà dyà mwoyi dyà mucìma
- dìsaamà dyà manungu

- d̄isaamà dyà m̄enu
- d̄isaamà dyà munda
- d̄isaamà dyà tulù
- d̄isaamà dyà dicì / d̄isaamà dyà macì
- d̄isaamà dyà diboba
- d̄isaamà dyà mashi / d̄isaamà dyà mashi àjika
- dibùùkà dyà mashi
- ditùùlà dyà difù
- dijìmina dyà lungènyì
- d̄ilamuka dyà kamònyi
- dyela dyà munda
- dipweka dyà kalisyumù / dipweka dyà kàlisyumù
- lukosù lwà cyàdi
- lukosù lwà mfwènka
- lwendu lwà mashi
- lukosù lwà mpùta
- lukosù lwà ngondo ìsambòmbò
- musàndà wà bilààdi / misàndà yà bilààdi
- musàndà wà ankililà / misàndà yà ankililà
- musàndà wà àmìbà / misàndà àmìbà
- m̄èsù àà m̄potà
- bulààlù bwà cileledi
- bwanga bwà misàndà
- kapyà kàà bakishi
- nzùbu wa bavyeèlà
- nzùbu wa cyondopelu
- nzùbu wa dipandila
- mukàjì nè difù

Type 4: SN = N + SCon

Avec un Inf:

SCon= connecteur + SV

SV = V_{inf} + N

- bwanga bwà kushipa k̄ishi / tw̄ishi

Avec un PaV

- lukosù lwà lùpàcisha cyàdi

Type 5: SN = N + SLoc

* SCon= Con + Loc

- lumònu lwà pabw̄ipì

* SLoc = Loc + N

- mashi mu manungu
- twîshi mu mênyi
- mpùta ku bisùlusulu
- bîshi kumênu
- twîshi mu bongù
- mâyi ku bisùlusulu
- cyuwùjà mu difù

Type 6: SN = Loc+ Dét

* Loc + N

- ku bapanda
- ku bwanga
- ku balelà
- ku bâna
- ku babùlùka
- kwà mungàngà (mùngàngà)

* Con + SLoc + N

- byà pa muminu

* Loc + Con + SN

SN = Con + N + PaV

- kwinshì kwà difù kùsaama

* N + Loc + Adj

- mukàjì kuulu kula

Type 7: SN = N + Adj

- mashi makesà

Type 8: SN = N + PaNA

- mênu àsaama
- mutù ùsaama

Type 9: SN = N + PaNA

- makàsà muwùla

Type 10: SN = N + N

- munda mpulumuku
- munda mashi
- mubidi kapyà (luuyà/mashika)
- kampala dibàkà
- cyuwùjà munda

2. Les formes verbales

2.1 Les infinitifs

- kwondapa / dyondapa
- kuditàngidisha
- kuditeetesha
- kufùùlulula
- kulènga
- kuumusha
- kupanda
- kuteeta
- kutùùla
- kudilùnga

2.2 Les syntagmes verbaux

2.2.1 SV à noyau gérondif

Type 1: Gérondif (verbal) + N

- dijika mashi
- dipeta shokà
- dipangila lupeepèlè
- dijandula bubèèdì / bubèdì
- dyela munda
- dishipa twîshi

Type 2: Gérondif + SN

* SN = N + SP

SP = Loc + N

- dipeta shokà ku mutù

Gérondif + SN

SN = N + SCon

SCon = Con + SLoc

SLoc = Loc + N

- dikòka mâyi àà mu mwongu

Type 3: Gérondif + SV

* SV = V_{inf} + Adv

- dibènga kumòna kula

2.2.2. SV à noyau infinitif

* Inf + N/SN

- kutùùla difù

- kwela mâyi
- kubàndisha mubidi
- kupangila (kwà) mupùùyà (lupeepèlè) / dipangila mupùùyà
- kufwà cipùùkà
- kwela cipimu
- kupima mubidi
- kupeta shokà
- kwangata mashi àà ku munu (SN = N + SP
SP = Con + Loc + N)

* Inf + Con + N

- kutàngila nè kaamù
- kumòngana nè mungàngà

* Inf + SAdjoint

SAdjoint = Conj + SV

SV = V_{inf} + N

- kupanda bwà kusabula mwâna

3. Les énoncés

* P à base prédicative

- mukùjì musòmba mu tròmpà

* P à base PaNA

- lwêpù lukèèpèla mu mubidi

* P à base verbale active

- dicì dîsaamisha mbàngà

*P à PaV

- mutù wèla mikenyi

Tableau synthétique

1	Formes nominales	137	Substantifs	57
			Gérondifs	11
			Emprunts	11
			Syntagmes nominaux	58
2	Formes verbales	32	Infinitifs	10
			Syntagmes verbaux	22
3	Les énoncés	4	Énoncé à base prédicative	1
			Énoncé à base PaNPa	1
			Énoncé à base PaV	2
4	Totaux	173		173

ANNEXE 2: Corpus: les termes français–cilubà

1. ablation: dyumusha, kuumusha
2. alcool: àlùkolà
3. amibiase: misàndà yà àmìbà / musàndà wà àmìbà
4. anémie: dijika dyà mashi, mashi makesà, dîsaamà dyà mashi, dibùùkà dyà mashi
5. angine: byà pa muminu, cilundù
6. antibiotique: bwanga bwà kushipa (kushebeya) twîshi
7. appendicite: kalwalwalwa, àpàndisità
8. arthrose: dîsaamà dyà manungu
9. asphyxie: dipangila mupùùyà / mupùùyà mupàta
10. asthme: mfwènka, lukosù lwà mfwènka
11. atteinte rénale: tonyi tupàtà/è
12. ausculter: kuteeta nè kaamù / kutàngila nè kaamù
13. avortement: ditùùla difù, kutùùla difù, kutùùla
14. bactérie: kîshi
15. bilharziose: bilààdì
16. blennorragie: mäsàsà
17. bronchite: lukosù lwà lùpàcisha cyâdi
18. broncho-pneumonie: lukosù lwà mpùta, mpùta ku bisùlusulu
19. cancer: kànsêrà / kànsêlà
20. cardiopathie: dîsaamà dyà mwoyi, mwoyi muwùla, dîsaamà dyà myoyi
21. carie dentaire: bîshi ku mênù, mênù àsaama, dîsaamà dyà mênù
22. cataracte: lusòngà
23. centre de santé: nzùbu wa cyondopelu, dîspensêrà
24. céphalée: mutù, mutù ùsaama, mutù wèla mikenyi
25. césarienne: dipanda / kupanda bwà kusabula mwâna
26. chancre dur: tumbulu
27. chirurgie (à la chirurgie): ku bapanda
28. choc: dijimija lungènyi, kufwà cipùùkà, kupeta shokà
29. choléra: munda mpulumuku, nkòlela
30. cirrhose: mucìma
31. clinique: kîlinikà
32. coma: cyoyoyi, dilènga, cyonà, dituuta cyonà, kulènga
33. comprimé: kìninà
34. conjonctivite: mèsù àà mpotà / mèsù àà mpòcyà, àpòlo
35. consultation: kuditàngidisha, kutàngila, kuteeta, kumòngana nè mungàngà, kuditeetesha
36. consultation prénatale: cipimu
37. convulsion: nkooyi, mpaalù, dyuminyina
38. coqueluche: lukosù lwà ngondu isambòmbù
39. crise: krîzà, dikùluka

40. décollement de la rétine: dilamuka dyà kamònyì
41. diabète: siikrà, dyàbètà
42. diagnostic: dijandula bubèdì
43. diarrhée: munda, dîsaamà dyà munda, dyela (dyà) munda
44. dispensaire: dispensêrà, ku bwanga, kwà mungàngà
45. docteur: mungàngà, doktêrà, ngàngàbukà
46. dysenterie: munda mashi
47. dysménorrhée: kwinshì kwà difù kùsaama
48. éclampsie: makàsà muwùla
49. éjaculer: kwela mâyi
50. éléphantiasis: buzevu
51. épiglottite: kakèlà
52. épilepsie: ciseki, mpaalù
53. épistaxis: milubàlubà, mashi àpàtuka mu dyùlu
54. examen: diteeta
55. femme enceinte: mukàjì nè difù, mukàjì kuulu kula, mukàjì mujìngà-kàne
56. fiche: fiishì
57. fièvre: mubidi kapyà / mubidi luuyà / mubidi mashìka, / kubàndisha mubidi
58. filariose: twîshi twà mu mashi, filêrà
59. gale: mpùsù, cikupi
60. gastrite: cifù
61. grossesse extra-utérine: mukùjì musômba mu trômpà, difù mu trômpà
62. hémarthrose: mashi mu manungu
63. hémorroïde: lusundu
64. hépatite: dîsaamà dyà mucìma, mucìma muwùla
65. hôpital: lùpitàadì, nzùbu wa cyondopelu, òpìtalà
66. hospitaliser: kubwela mu lùpitàadì / kubwela mu òpìtalà
67. hypertension: lwendu lwà mashi
68. hypocalcémie: dipweka dyà kàlisyumù mu mashi, lwêpu lukèèpèla mu mubidi
69. hypoglycémie: siikrà, dyàbètà
70. hypotension: lwendu lwà mashi
71. infection urinaire: twîshi twà mu mênysi, mìkrôbà
72. infirmier: mungàngà, ngàngàbukà
73. injecter: kutwà lushìngì
74. intoxication: kudilùnga, mulùngù, kunwà bwanga bupìta, dilèwuka
75. jaunisse, ictère: jòdnisà, jòdnà
76. kinésithérapie: siyânsà, kine
77. kwashiorkor: diboba, nsàdi, kwashì
78. laborantin: mungàngà wa ku làbo
79. laboratoire: làbo
80. laryngite: byà pa muminu

81. leucémie: kansêrà kàà mu mashi
82. liquide amniotique: mâyi
83. lit gynécologique: bulààlù, bulààlù bwà cileledi
84. maladie du sommeil: lùbùngù, dîsaamà dyà tulù
85. malaria: màlèyalèya, twîshì twà mu mashi, màlàriya
86. malaria cérébrale: katombùtombù, bwongù bùyùnguluka, mukupa, dibùluka
87. malnutrition: diboba, dîsaamà dyà diboba, nsàdi
88. maternité (à la maternité): ku balelà, nzùbù wa bavyèla, màtrèmìte
89. matrice: mààtrisà
90. méningite: menènjiità, twîshi mu bwongù
91. microbe: kîshi
92. myopie: dibènga kumòna kula, lumònu lwà pabwîpi
93. neurologie: ku babùlùka
94. opération chirurgicale: dipanda, dipità ku keelà
95. opérer: kupanda
96. orgelet: kampala dibàkà
97. otite: dîsaamà dyà dicì (macì), minsonjì / misònji
98. otite mastoïde: dîsaamà dyà dicì (macì), dicì dîsaamisha mbàngà
99. otite purulente: tufina mu macì, minsènsa mu macì / minsònji
100. pédiatrie: ku bâna
101. perforation intestinale: mala matùbùùka
102. perfusion: kwela mâyi
103. photographie: fòòto
104. plaie: mpùta
105. plâtre: cùto
106. pleurésie: mâyi ku bisùlusulu
107. pneumonie: mwoyi muwùla
108. poche des eaux: kalòwa
109. poliomyélite: tunèka
110. ponction lombaire: dikòka mâyi àà mu mwongu
111. prélever la température: kwela cipimu
112. presbytie: bupofu
113. prostate: pròstatà
114. réanimation: kufùùlulula
115. respiration artificielle: mbòmbonà
116. rhumatisme: binyòkà, ngàlàbàtizà
117. rougeole: kàntembela
118. salle d'accouchement: cileledi
119. salle d'opération: nzùbù wa dipandila, cipandidi
120. schistosomiase: misàndà yà bilààdì
121. séringue: lushìngì
122. sida: kashèlèlèkà, sîda, bùdì bwendà
123. sinusite: sìnizità

-
124. soluté: mâyi
125. spermatozoïde: mâyi àà balùmà, bilùmà
126. stérilisation: dishipa twîshi
127. stéthoscope: kaamù kàà mu macì
128. strongyloïde: misàndà yà ànkililà
129. syphilis: kàsèndà
130. tétanos: ntèntànosà
131. toux: lukosù
132. traitement: londàpu, dyondapa
133. traumatisme: dipeta shokà, dipèta ku
134. traumatisme cranien: dipeta shokà ku mutù, dipeta ku ku mutù
135. trompe: trômpà
136. trouble mental: ditomboka, katombùtombù, dipaala
137. tuberculose: cyâdi, lukosù lwà tashi / cyâdi
138. tumeur abdominale: cyuwùjà munda, cyuwùjà mu difù
139. typhoïde: cìfòyîdà
140. ulcère de l'estomac: mpùta mucifù
141. urétrite gonococcique: mäsàsà
142. vaccin: cisàlu
143. varicelle: makùbùngù
144. variole: mbàlàngà
145. ver: musàndà
146. vermifuge: bwanga bwà misàndà
147. verminose: musàndà
148. video: vidèò
149. virus: kîshi
150. zona: kapyà / kadilu kàà bakîshi

Kiswahili Verbs: A Lexicographical Challenge

Z.s.m. Mochiwa, *Department of Kiswahili, The Open University of Tanzania,
Dar Es Salaam, Tanzania (zasamociwa@yahoo.com)*

Abstract: This article makes three significant claims about Kiswahili verbs. By investigating the verbal root *pend-* 'love' the article claims that the agglutinative nature of Bantu languages is at the core of the morphological fecundity of Kiswahili verbs. Evidenced both vertically and horizontally, the fecundity brings in, respectively, extensions and derivations. The article claims further that each of the extensions of the verb can, theoretically at least, participate in derivational processes. The second claim is that this fecundity triggers lexicographical problems of choice and semantic analysis. Specifically, the article underscores the need for selectivity of information to control the density of the dictionary entry. Yet, selectivity presupposes a thorough analysis of the morpho-syntactic behaviour of the verb in its many extended and derived forms. Indeed, such an analysis presupposes, in turn, an eclectic use of many linguistic theories. Bantu lexicography must be up front in making theoretically sound decisions. Finally, on the basis of linguistic theories, lexicography becomes a concretization of the mental lexicon claimed to be part of the competence of the native speaker.

Keywords: KISWAHILI, VERBS, LEXICOGRAPHY, LEXICON, DICTIONARY, INFLECTION, EXTENSION, DERIVATION, CAUSATIVE, SOUND-MEANING, ROOT, CITATION

Opsomming: Kiswahiliwerkwoorde: 'n Leksikografiese uitdaging. Hierdie artikel maak drie betekenisvolle bewerings oor Kiswahiliwerkwoorde. Deur 'n ondersoek van die verbale kern *pend-* "liefhê" beweer die artikel dat die agglutinerende aard van die Bantoetale aan die kern lê van die morfologiese vrugbaarheid van Kiswahiliwerkwoorde. Soos sowel vertikaal as horisontaal getoon, lei die vrugbaarheid tot suffiksering en afleiding onderskeidelik. Die artikel beweer verder dat elk van die agtervoegsels van die werkwoord, teoreties altans, aan die afleidingsprosesse kan deelneem. Die tweede bewering is dat hierdie vrugbaarheid leksikografiese probleme van keuse en semantiese ontleding veroorsaak. Die artikel onderstreep veral die noodsaaklikheid van die selektiwiteit van inligting om die digtheid van die woordeboekinskrywing te beheer. Selektiwiteit veronderstel egter 'n deeglike ontleding van die morfosintaktiese gedrag van die werkwoord in sy talle gesuffikseerde en afgeleide vorme. Op sy beurt veronderstel so 'n ontleding inderdaad 'n eklektiese gebruik van talle taalkundige teorieë. Bantoeleksikografie moet die eerste wees in die maak van teoreties gegronde besluite. Op grond van linguistiese teorieë word die leksikografie uiteindelik 'n konkretisering van die mentale leksikon wat na bewering deel van die vermoë van die moedertaalspreker is.

Slutelwoorde: KISWAHILIWERKWOORDE, LEKSIKOGRAFIE, LEKSIKON, WOORDEBOEK, VERBUIGING, SUFFIKSERING, OORSAAKLIK, KLANK-BETEKENIS, KERN, AANHALING

1. Introduction

A dictionary is the result of an all-encompassing sociolinguistic research. According to Gleason (1967: 101), a dictionary is 'the meeting place of all the systems, linguistic and non-linguistic which bear relevantly on speech behavior'. The linguistic side of the project reveals all, or, at least, as many as possible of, the idiosyncrasies of the language system of which the dictionary is compiled. In the light of this, a dictionary must enable the user to come face to face with the arbitrary nature of the language. Put somewhat succinctly, a dictionary reveals the non-correspondence between form and meaning in language.

Looked at from this same angle, a dictionary may be considered as being a near mirror image of the lexicon claimed to be part of a native speaker's competence. According to generative grammarians, this lexicon in the native speaker's mind contains a list of morphemes whose meanings cannot be predicted by mere inspection of their forms. One is tempted to add here that the dictionary is in many respects a concretization of the mental lexicon native speakers are claimed to have. Representing this school of thought, Kenstowicz and Kisseberth (1979: 3) state:

In generative grammar it is assumed that the morphemes of a language are stored by the speaker in a special listlike device called a *lexicon* or *dictionary*, which contains all of the truly unpredictable, idiosyncratic information about the behavior — syntactic, semantic, phonological — of each morpheme known to the speaker.

Although a parallel can be drawn between the dictionary and the lexicon, the latter is much more detailed than the former. This means that the compilation of the dictionary goes hand in hand with some deliberate selectivity. Bearing in mind the anticipated user's needs of the dictionary being compiled, lexicographers are forced to select certain types of information. What is more, considerations outside the field have considerable influence on lexicographers' work. These include usability of the product, cost issues and possibly the size of the finished product. These important considerations made Zgusta (1971: 16, 17) sound the following warning to lexicographers:

The lexicographer is doing scientific work, but ... he publishes it for users whose pursuits are always more practical, at least as regarded from his own point of view ... In other words, the user of the dictionary does not wish, at least usually, to have the purely lexicographic problems presented, but to have them solved.

As has already been pointed out, the density of information included in a dictionary varies, depending on the anticipated user. A learner's dictionary will be different from one for the use of a native speaker. The lexicographer rightly assumes that the user, being a native speaker of the target language, has acquired an exhaustive grammatical competence. The user's grammatical competence, for example, obviates the need for recording some of the morphological processes characteristic of the language. To the learner, however, the dic-

tionary is a tool for learning the target language. It must contain information that will enable him/her to acquire competence in it.

In their bid to control the density of information to be recorded in their dictionaries, Bantu lexicographers are, time and time again, uncertain of what to include. As a result they tend to lay undue emphasis on matters of frequency regardless of the target user. In accordance with the extent to which users try to acquire the target language exhaustively, lexicographers must ensure that the morphology of the most frequently used words are given. The overall success of lexicographers' products depends to a large extent on what they choose to record.

One is, however, apt to point out that the choice of what to include must be such that the given information enables the user to acquire an accurate account of the morphological structure of the language. This is further emphasized by Prinsloo and De Schryver (2001: 188) who claim, among others:

The basic aim of the lexicographer is to guide the user in respect of the properties/features/characteristics/use/meaning of the lemma, i.e., to know the word.

To be more precise, a dictionary representing the native speaker's lexicon must accurately characterize both the inflections and derivations found in the language being described.

2. The problem

The agglutinating characteristic of Bantu languages affects verbs more significantly than they do other categories. Verbs change morphologically as they receive affixes either horizontally or vertically. They receive affixes horizontally during extensions or, more precisely, during their inflections. Kiswahili, for example, has, according to Kiango (2000: 104), eight verbal affixes¹. Among these, three are less productive while the other five are very much so.

The five productive affixes present formidable problems to the lexicographer in two ways. On the one hand, a single root of a verb can use a given affix twice or even more. On the other hand, each root that has acquired a suffix can, at least theoretically, participate in the vertical morphological process of derivation². This means, any root that acquires an affix can nominalize or deverbalize into an adjective.

The theoretical implication of this behaviour is that any verb in these languages is a 'word factory', so to speak. Many scholars, including Vitale (1981), Bokamba (1981), Keach (1985), and Prinsloo and De Schryver (2001), among many others, account for this state of affairs by appealing to the morphology with which Kiswahili is endowed. All of them claim that its complex morphology is entwined with its syntax. According to Givón (1971), the complexity is a result of the loss of syntactic features which forced otherwise free morphemes to be bound.

We are apt to ask: How can a Kiswahili verb then be exhaustively treated

in a dictionary? Put somewhat differently: Can the lexicographer record and analyze each and every form that derives from the simple root? Is it possible to use a single verb entry for all its extensions as well as its derivatives such as nouns and adjectives? In other words: Can the fecundity of a Kiswahili verbal root such as *pend-* 'love' fit into a single dictionary entry? If the answer to this question is "no", how does the lexicographer identify his/her lemmata? De Schryver and Prinsloo (2002) remind lexicographers that, if the modern Bantu dictionary is to be really practical and useful, all verbs and their derivations likely to be looked up need to be included.

The purpose of this article is to explore the morphological and semantic fecundity of *pend-* 'love', a simple verb root in Kiswahili. By looking at its horizontal as well as its vertical changes in the acquisition of affixes, our main focus will be on two important issues. On the one hand, we intend to show the morphological complexity of the forms, both verbal and deverbal, resulting from the affixation process of the root. On the other hand, we want to study the semantics of the resulting verbal forms.

The study is carried out in two ways. First, we study the morphosyntactic patterning of the suffixes. Issues such as which suffix precedes another and whether or not a single suffix can be used more than once will be within the interests of the article. Second, we study the type of meanings the resulting verbal forms express. Specifically, we want to establish whether or not the verbal forms express compositional meanings. Before we embark on this issue, we need to articulate a few, but fundamental, assumptions.

3. Theoretical underpinnings

A dictionary has been described as a repository of idiosyncrasies (Kenstowicz and Kisseberth 1979). These idiosyncrasies explain the arbitrary nature of language in general. The most widely acknowledged of these idiosyncrasies is the form–meaning relation. Meanings of words are, by and large, unpredictable. We cannot guess the meaning of a given word by inspecting its linguistic form. Thus, a dictionary is intended to reveal to the user, both the semantic and morphosyntactic peculiarities of words found in the language for which it is compiled.

Languages endowed with a complex morphology, as Bantu languages are, trigger many questions. The most problematic of these in Kiswahili lexicographical projects is the identification of citation forms to be entered in the dictionary. How many extensions deserve entries? How many of the derived forms must be treated as separate entries? If a single citation form is identified for a verbal root like *pend-* 'love' and the rest is left to the user, can he/she, being a learner, find his/her way through the linguistic maze?

Kiango (2000: 21) who notes this problem, argues that it is compounded by Western theories of lexicography. His argument is that lexicographers working on Bantu languages need to be careful with the use of theories which arose

from typologically different languages. He observes the following:

Lexicographical theories and methods are applied to natural languages which by their nature, have different phonological, morphological and syntactical structures.

Although dictionaries differ depending on specific factors, most of which lie outside the lexicographical field, lexicographers exclude information said to be derivable from the grammatical competence of the user. In Kiswahili, for example, once *cheka* 'laugh' is semantically analyzed, one can 'safely' assume that the user can derive *chekeka* 'be laughable'.

It is, in this vein, that *Kamusi ya Kiswahili Sanifu*, 'The Standard Kiswahili Dictionary', henceforth KKS, cites a few verbal extensions after the treatment of each verb. However, this is fine as long as the dictionary is for the use of native speakers. Suppose now that the dictionary is for language learners³. They use a dictionary to learn both the morphosyntactic and semantic features of the target language.

Compiling a Kiswahili dictionary is a morphosyntactic project. The lexicographer must simultaneously and aptly deal with the morphology of the language, which, needless to say, is not only intricate but also triggers syntactic issues. When it is extended, a verb, for example, acquires new syntactic features, which affect the meaning it expresses. It is, as this article wants to argue, misleading to assume that the user will arrive at the forms omitted.

Although the majority of the languages are yet to be fully analyzed, lexicographers do not take much trouble in their study of the forms they enter into the dictionaries. More often than not, the user who is at the 'mercy' of the dictionary, is given words falsely claimed to be variants of each other. Yet, a close look at them reveals that each one is derived from a different verbal root. KKS, for example, claims that *mpotovu* and *mpotevu* are variants of each other.

The morphological histories of the two nouns reveal clearly that they are derivatives of two different verbs. The noun *mpotovu* is derived from the verb *potoa* 'corrupt' whereas *mpotevu* is derived from *potea* 'be lost'. Obviously, the two derived forms cannot express the same meaning. *Mpotovu* is 'someone who corrupts people' whereas *mpotevu* is 'a prodigal person/someone who tends to disappear'. *Mpotevu* also possesses the meaning of 'squanderer'.

That the two derivative nouns are treated as variants is indicative of the dearth of morphological analysis of the language. Indeed, this is like assuming on the lexicographers' part that the morphology of the target language can, and indeed does, produce synonyms. This is again, as this article wants to argue, a misconception. Lexicographers who have not had both rigorous and vigorous training in theoretical linguistics lack the apparatus needed to reveal the mysteries of words.

What this discussion aims to show is that lexicographers must, first and foremost, be theoretical linguists. Such lexicographers have the skills needed for undertaking the project. Given their training, they can resolve issues of a theoretical nature. According to Zgusta (1971: 19), 'some of the problems of

lexicographic analysis and presentation are surprisingly general, irrespective of the language in question'.

Zgusta's observation is crucial despite Kiango's concern about the use of 'imported' theories. Zgusta underscores the theoretical fact that, at a certain level of abstraction, human languages are similar, if not identical altogether. Theories that conceptualize language at that level are the ones the lexicographer needs. Kiango's observation is equally valid, because the lexicographer needs language-specific parameters (cf. Radford 1997). In other words, both Zgusta's and Kiango's points of view are important for the lexicographer to take seriously.

The theoretical linguist is in a better position to conceptualize the nature of language. He/she is in an informed position to deal with theoretical issues. A proper theoretical training makes the lexicographer adept at discovering the rule-governed nature of language.

The theoretical training emphasized here is a fundamental knowledge of the nature of language that guides the lexicographer in making the right decisions. The lexicographer is not supposed to display his/her linguistic expertise in the dictionary he/she is compiling. His/her linguistic expertise enables him/her to accurately characterize the idiosyncracies of the language.

Firmly grounded in theoretical linguistics, lexicographers would not confuse *mpotovu* and *mpotevu*. An extension such as *chemkika* purported to trigger from *chemka* 'boil' is another indication of theoretical misconceptions. With a good theoretical grounding, these misconceptions could be minimized to avoid the mistakes cited above which have, undoubtedly, far-reaching consequences for the learner.

A lexicographical theory is much broader than any ordinary linguistic theory. It covers, according to Zgusta (1971: 15), 'the whole structure of the language in question' and 'the culture of the respective community in all its aspects'. In the first issue of *Leksikografičeskij sbornik bratislavský*, he states:

The theory of lexicography is connected with all the disciplines which study the lexical system: semantics, lexicology, grammar, stylistics.

Knowing the importance of language theories, Weinreich (1967: 26) had occasion to lament 'the indifference lexicography displays towards its own methodology'. Clear theoretical and methodological apparatus is a good roadmap to lexicographical success. With this in mind, we can begin to investigate the verbal root *pend-* as it acquires suffixes. We will allow the root to acquire one suffix at a time.

4. Extension suffixes

The number of extension suffixes identified by Kiango (2000) differs from that of Khamis (1972) who claims that Kiswahili has eleven. According to Khamis, the suffixes are the mood, i.e. the vowel suffix (*vs*), the conversive/reversive

suffix (crs), the intensive suffix (is), the static suffix (sts), the durative suffix (ds), the potential suffix (pos)⁴, the stative suffix (ss), the applicative (sometimes referred to as the prepositional/applied) suffix (as), the causative suffix (cs), the passive suffix (pas) and the associative/reciprocal suffix (rs).

Except for the conversive/reversive and the intensive, all the suffixes change both the morphosyntactic and semantic patterning of the resulting verbal forms. The conversive/reversive and the intensive suffixes change the semantics of the verbal forms but the morphosyntactic features remain the same.

When acquired by the root, the conversive/reversive suffix, for example, realized as *-u-* or *-o-*, makes it express an antonymic meaning. It changes the meaning from 'do' to 'undo'. Thus, for example, *funga* 'close', *panga* 'arrange' and *choma* 'pierce' respectively become *fungua* 'untie/open', *pangua* 'disarrange' and *chomoa* 'unpierce'.

The intensive suffix, on the other hand, adds vigour and/or continuity as semantic features to the action of the verb. Without it, the verbal roots *pig-*, *chek-* and *end-* respectively mean 'hit', 'laugh' and 'go'. On acquisition of the intensive suffix, however, the new verbal forms *pigilia*, *chekelea* and *endelea* respectively mean 'hit hard continuously', 'laugh continuously/repeatedly', 'go unabated/continue without stopping'. Apart from these semantic features added to them, the verbal forms acquire no syntactic features. Intransitive verbs, for example, remain intransitive and transitive verbs remain transitive.

For purposes of this article, we investigate five suffixes, namely, the applicative, the causative, the stative, the intensive and the associative/reciprocal. What we intend to establish is their morphosyntactic patterning as well as their semantic input. At the same time, we intend to find out whether or not, on acquisition of the suffixes, the resulting verbal forms express compositional meanings. This will specifically be our major focus when the verbal forms result from the acquisition of more than one suffix. To begin with, we follow the root *pend-* as it acquires one suffix at a time.

4.1 One extension suffix

The verbal root *pend-* attracts all the suffixes, except the conversive/reversive, the durative, and the static. That *pend-* cannot acquire these suffixes is exemplified by the unacceptability of the following verbal forms, each of which uses the root.

- | | | |
|-----|---------------------|--|
| (1) | <i>penda</i> 'love' | * <i>pendoa/pendua</i> 'unlove' (conversive) |
| | <i>penda</i> 'love' | * <i>pendaa</i> 'love all over' (durative) |
| | <i>penda</i> 'love' | * <i>pendama</i> 'stay in love' (?) (static) |

Other verbal roots, however, can acquire these suffixes as can be seen from the following:

- | | | |
|-----|----------------------|---|
| (2) | <i>funga</i> 'close' | <i>fungua</i> 'open/tie – untie' (conversive) |
| | <i>paka</i> 'smear' | <i>pakaa</i> 'smear all over' (durative) |
| | <i>tua</i> 'perch' | <i>tuama</i> 'perch and stay' (static) |

When the root *pend-* is allowed to acquire a suffix, the following verbal forms result:

- (3) *pend-a* 'love/like'
 root + vs
pend-an-a 'love/like each other'
 root + rs + vs
pend-e-a 'love/like for/on behalf of'
 root + as + vs
pend-ek-a 'be lovable/likable'
 root + ss + vs
pend-esh-a 'cause to be lovable/likable'
 root + cs + vs

All the verbal forms under (3) above result after the root has acquired a certain suffix in addition to the vowel suffix. The first verbal form, however, is somewhat different from the others, because the root has acquired the vowel suffix only. The last verbal form does not seem to result from a direct acquisition of the causative suffix. It seems much more plausible and theoretically tenable to argue for the acquisition of the stative suffix prior to the causative suffix. Two reasons are adduced in support of this claim. On the one hand, the meaning expressed by the verbal form *pendesha* fits into the semantic frame 'cause to be verbable' rather than 'cause to verb'.

Thus, *pendesha* shares the same semantic frame with verbs derived from adjectives such as the following:

- (4) *fupi* 'short'
kamili 'complete'
bora 'better'
refu 'long'

Such adjectives become verbs after the acquisition of the stative suffix thus:

- (5) *fupi-k-a* 'become short/shorten'
kamili-k-a 'become complete'
bor-ek-a 'become refurbished/better than before'

Like *pendeka* 'become lovable', the stative verbal forms under (5) can acquire the causative suffix *-sh-* as under (6).

- (6) *fupisha* 'cause to become short'
kamilisha 'cause to become complete'
boresha 'cause to become better than'

The causative suffix *-sh-* expresses the meaning in which someone or something triggers causation of and/or capability in which someone or something becomes verbed. The causative *sh* is specifically found with verbal forms that have acquired the stative suffix. Verbal forms that use *sh* for the causative fit into the semantic frame 'be verbable/become verbable'.

The claim that the introduction of the causative suffix *-sh-* is indirect is supportable on phonological grounds. Verbal forms which take it are those which, when analyzed, have their roots ending in the velar stop /k/ and, occasionally, the alveolar stop /t/. Such verbal forms almost invariably take the *-sh-*. Here are a few examples in addition to the above:

- | | | |
|-----|--|---|
| (7) | <i>pita</i> 'pass' | <i>pisha</i> 'cause to pass' |
| | <i>pata</i> 'get' | <i>pasha</i> 'cause to get' |
| | <i>choka</i> 'tire' | <i>chosha</i> 'cause to tire' |
| | <i>ruka</i> 'jump/fly' | <i>rusha</i> 'cause to fly/jump' |
| | <i>hangaika</i> 'be unsettled/startled/
troubled' | <i>hangaisha</i> 'cause to be unsettled/star-
tled/troubled' |

Although the article will not go into the details of the morphophonology involved, what can easily be seen happening is that the velar stop is replaced by the causative. Thus, the conclusion that can be drawn here is that verbal roots ending in /k/ or /t/ have their causative suffix in the form of a voiceless palatal fricative *-sh-*.

Some of the verbal forms with roots ending in the velar or alveolar stop are found to have two causative verbal forms or one but which is acquired after the stative. Without the stative suffix, the causative suffix derives the following verbal forms:

- | | | |
|---------|----------------------|---|
| (8) (a) | <i>pita</i> 'pass' | <i>pisha</i> 'cause to be passable' |
| | <i>pika</i> 'cook' | * <i>pisha</i> 'cause to be cooking/cooked' |
| | <i>cheka</i> 'laugh' | * <i>chesha</i> 'cause to be laughable' |

As can be seen, *pisha* as the causative of both *pika* and *pita* is not current in the language. This is partly due to the morphological clash in which the causative is forced to serve both verbs. If, however, the three verbs acquire their causative after their stative, the verbal forms which result are, without exception, current as becomes evident from the following set of data:

- | | | | |
|---------|--------------|----------------|---|
| (8) (b) | <i>pita</i> | <i>pikita</i> | <i>itisha</i> 'cause to be passable' |
| | <i>pika</i> | <i>pikika</i> | <i>pikisha</i> 'cause to be cooking/cooked' |
| | <i>cheka</i> | <i>chekika</i> | <i>chekisha</i> 'cause to be laughable' |

Thus, both *pisha* and *pitisha* are current in the language. The former means 'cause to pass', whereas the latter means 'cause to be passable'. Contrariwise, *chesha* does not seem to be current, although *chekisha* is indeed current. The stem *-cheshi*, however, is current because it is found in the noun *mcheshi* which means 'a jovial person'.

When the verbal root acquires the suffix, the resulting verbal form has the semantic frame 'cause to verb'. If, however, the causative is introduced after the stative suffix, the resulting verbal form means 'to cause to be verbed or verbable'. Thus, whereas *pisha* means 'cause to pass', *pitisha* means 'to cause to be passable'. The fact that *pisha* 'cause to cook' does not exist, is probably due to

a possible clash between *pisha* from *pita* 'pass' and the causative of *pika* (*pisha*). Thus, *pik-ik-* 'be cookable' instead of *pik* 'cook' is allowed to causativize.

An immediate counterexample to the above claim is the verbal root *f-* 'die'. This root extends to *fisha* 'cause to die'. It cannot be argued that *fa* 'die' extends to *f-ik-a* which in turn, with the introduction of the causative suffix, extends to *fisha*. Other verbal forms, which, *prima facie*, look like possible counterexamples, are given under (9):

- (9) *kumbuka* 'remember'
anguka 'fall down'
kurupuka 'dart away'

The three forms, however, are not counterexamples. All of them are essentially stative verbs. For the form *kumbuka*, for example, the existence of *kumbukumbu* 'memory' does suggest in a strong way that the true root of *kumbuka* 'remember' is not *kumbuk-*, rather *kumbu-* and the velar consonant /k/ was introduced by a stative rule. The other two roots extend as follows:

- (10) (a) *angu-a* 'cause to fall'
angu-k-a 'become falling'
angu-sh-a 'cause to become falling'
- (b) *kurupu-a* 'startle/cause to dart away'
kurupu-k-a 'be in the darting/startled'
kurupu-sh-a 'cause to be in the darting away'

The two examples above give more evidence in support of the claim that *-sh-* is a causative suffix specific to stative verbs. The semantic contrast between *angua* and *angusha* lends support to the claim that the morphological engine of a language does not produce synonyms. A study of *angua* shows that it is inherently a causative verb expressing the meaning 'make to fall' or 'verb the object'.

Likewise, *kurupua* has the same semantic content as *angua*, namely, 'make to dart away'. When these verbs acquire the overt causative, they change their semantic frame from 'verb the object' to 'cause the object to be verbable'. Thus, there is no synonymy between *angua* and *angusha* or *kurupua* and *kurupusha*. This morphological behaviour is characteristic of many other verbal forms. Verbal forms such as those under (11) exemplify this.

- (11) (a) *ondoa* 'cause to move/remove'
ondoka 'be in a state of moving'
ondosha 'cause to be in a state of moving'
- (b) *opoa* 'cause to be safe/save'
opoka 'be in a state of safety'
oposha 'cause to be in a state of safety'

The verbal form *dondoa* 'pick up' allows the formation of *dondosha* but not *dondoza*, as we would expect. We are given to understand that *dondoa* does not

participate in the causative suffix introduction rule. Indeed, if it were to apply, the rule would produce *dondoza* which means 'make someone verb'. *Dondoza* is semantically different from *dondosha*⁵ which means 'cause to become dropping'.

The claim here is that the causative meaning inherent in the root itself is differentiated from the one introduced by the morphological rule. The latter comes in to express 'a state of becoming verbed' whereas the former, the inherent one, expresses the meaning 'cause to verb'.

We might add here that the inherent causative in *ondoa* expresses an 'actor' meaning. The introduced causative suffix, on the other hand, expresses an 'acted upon' meaning. It is plausible to claim that the difference between the two is that the inherent causative focuses on the subject of the verb, whereas the other focuses on the object. For the sake of clarification, let us study the following sentences in which the two verbs *angua* and *kurupua* are in use:

- (12) *Mosi anaangua maembe* 'Mosi is making the mangoes fall'
Maembe yanaanguka maembe 'Mangoes are falling down'
Mosi amekurupua sungura 'Mosi has startled the rabbit'
Sungura amekurupuka 'The rabbit has darted away'
Mosi anaangusha maembe 'Mosi is causing the mangoes to be falling'
Mosi anakurupusha sungura 'Mosi is causing the rabbit to dart away'

All these examples support the claim that the stative suffix introduction rule feeds the causative rule.⁶ What is also important here is the morphosyntactic patterning of the suffixes. In a dictionary showing possible extensions, it is misleading to indicate that the root *pend-* is immediately followed by a causative suffix of any kind. It is *pendeka* 'become/be lovable' which can receive a causative to produce the verbal form *pendesha* 'cause to be lovable'.

In other contexts, the causative suffix takes various phonetic forms. A verbal form ending in two vowels has the alveolar fricative /z/ for the causative suffix. The sequence of vowels comes about in one of two ways. On the one hand, if the verbal root ends in a vowel, then the vowel suffix becomes the second vowel. On the other hand, a verbal root which has acquired a stative suffix, has the vowel suffix at the end of the verbal form as its second vowel. Thus, on receiving the applicative suffix, *pend-* becomes *pendea* and the causative verbal form becomes *pendeza* which means 'cause to love on account of'. That this is indeed the case, is exemplified by the following set of data.

- (13) (a) *kaw-a* 'delay'
kaw-i-a 'delay someone on account of'
kaw-i-z-a 'cause to delay someone on account of'
- (b) *chag-a* 'work with renewed vigour'
chag-i-a 'work with renewed vigour on account of'
chag-i-z-a 'cause someone to work with renewed vigour on account of'

The introduction of the reciprocal suffix in (14) below makes the preverbal and postverbal nouns share subjecthood as well as objecthood. The reciprocal suffix

causes the action of the verb to be bidirectional in which the hitter is simultaneously hit. Let us have a look at the data under (14).

- (14) *Juma anapenda msichana* 'Juma loves a girl'
Juma na msichana wanapendana 'Juma and the girl love each other'
Juma anapendana na msichana 'Juma and the girl love each other'

Semantically, on the acquisition of the reciprocal suffix, both the subject and the object of the first sentences, namely *Juma* and *msichana*, do not only share subjecthood but also objecthood. An intransitive verbal root, however, cannot acquire the reciprocal suffix, because it has a preverbal noun only. Thus, this becomes a constraint when the reciprocal suffix introduces the sharing of subjecthood and objecthood. Likewise, the reciprocal suffix cannot be introduced to a stative verbal form for the same reason. The claim is corroborated by the set of data below in which intransitive verbs are used:

- (15) *chemka* 'boil'
sita 'hesitate'
lia 'cry'
tokota 'boil (said of heavy liquids)'

Introducing the reciprocal suffix to any of these verbal forms results in unacceptable verbal forms, as can be seen from the examples under (16) below:

- (16) **chemkana* 'boil each other'⁷
**sitana* 'hesitate each other'(?)
**liana* 'cry each other'(?)
**tokotana* 'boil each other'(?)

Any transitive verb, however, can acquire a reciprocal suffix. Indeed, an intransitive verb that has acquired a causative suffix or an applicative suffix, can acquire the reciprocal suffix. In the following set of data, simple verbal roots are detransitivized by the stative suffix and then retransitivized by the causative suffix:

- (17) *som-* 'read'
som-ek-a 'be readable'
som-esh-a 'cause to be readable'⁸

sem- 'say'
sem-ek-a 'be sayable'
sem-esh-a 'cause to be sayable'

imb- 'sing'
imb-ik-a 'be singable'
imb-ish-a 'cause to be singable'

The three verbal forms can convert into reciprocals as under (18).

- (18) *som-esh-an-a* 'make reading accessible to each other'

sem-esh-an-a 'make saying possible for each other'
imb-ishan-a 'make singing possible for each other'

By virtue of the applicative it has acquired, the third verbal form under (3) above, namely *pendea*, has become a ditransitive instead of a monotransitive verb. In other words, instead of a single postverbal noun, the verbal form has two postverbal nouns. This is readily exemplified by the set of data under (19).

- (19) *Juma anapenda msichana* 'Juma loves a girl'
Juma anampendea nini msichana yule? 'For what reason does Juma like the girl?'
Juma anampendea tabasamu zake 'Juma loves the girl on account of her smiles'

The applicative suffix is introduced on both transitive and intransitive verbal roots. With the introduction of the suffix, intransitive verbal forms become transitive whereas the transitive ones become ditransitive. Intransitive verbs such as the following become transitive following the introduction of the suffix:

- (20) *f-a* 'die' *f-i-a* 'die on account of/against the wish of'
lal-a 'sleep' *lal-i-a* 'sleep on/for/at'
ja-a 'be/become full' *ja-li-a* 'be/become full for the sake of'
ka-a 'sit' *ka-li-a* 'sit for/on behalf of/for the sake of/against the wish of'
fa-a 'be suitable' *fa-li-a* 'be suitable for/suit'

A detailed analysis of this type prevents the lexicographer from making wrong decisions. It is, for instance, misleading to indicate that every verbal root can directly receive a causative suffix in which the meaning remains to be 'cause to verb'. Or, for that matter, the user of the dictionary can be misled into believing that every verbal root can acquire a stative suffix, regardless of whether or not it is transitive. The lexicographical indications found immediately after verb entries suggest that each root can acquire a stative suffix. Yet this does not represent the morphological reality of the language. Intransitive verbs such as *gumi-a* 'roar', *ja-a* 'be full', and *ish-a* 'be finished/exhausted', are inherently stative. Allowing them to acquire the suffix results in verbal forms with double statives that are hard to interpret as can be seen from the following which are neither current nor potential:

- (21) *gumi-a* 'roar' **gum-ik-a*
ja-a 'be full' **jal-ik-a*⁹
ish-a 'be exhausted' **ish-ik-a*
po-a 'be cool/cold' **pol-ek-a*¹⁰

As far as their semantic patterns go, it can be claimed that verbal forms with single suffixes express compositional meanings. Thus, the meaning of a verbal form carrying a suffix will express the meaning of the root plus that of the suffix. It is however, important to note that, occasionally, a given verbal form may

carry more than one suffix. Thus, although it is *prima facie* a single root, a close analysis will reveal that it has another affix which is somewhat invisible as a result of morphophonemic processes it has undergone. In such a situation, the verbal form gives a false reading. The best examples are *ondoa* and *pendea*. Without prior knowledge of the verbal forms from which these are derived, chances are that the two causative forms are given the reading 'to verb' instead of 'to be verbed' or 'to be verbing'.

As has been pointed out earlier, each of the verbal forms can convert into a noun. The language has four nominalizing suffixes, namely *-o*, *-e/-w*, *-ji*, and *-i*. Each of these nominalizers has a specific meaning; *-o*, for example, expresses the result of the action of the verb, the process or the instrument of the verb. Let us study the following examples:

- (22) *pend-o* 'the result of the verbing, love'
pend-an-o 'the result of loving each other'
pend-e-o 'the cause/result of being lovable'
pend-ek-o 'the state of being lovable'
pend-e-sho 'the causal reason for being lovable'

Depending on idiosyncratic features of the root, derived nouns such as those under (22) acquire different noun class prefixes that somewhat modify their meanings. Thus, *upendo* and *mapendo* are different. *Upendo* is 'the feeling of love' whereas *mapendo* is 'all those acts which reveal the feeling of love'. Likewise, *upendano* is 'the feeling of loving each other' and *mapendano* are 'all the acts which reveal the feelings of people who love each other'.

The nominalizers *-i* and *-ji* denote the 'doer' of the verb, i.e. they represent the subject of the verb. The former bears an added meaning of a professional verber, a regular doer whereas *-ji* denotes occasionality, irregularity and lack of expertise and probably lack of seriousness. Thus, *mpenzi* is someone who loves seriously, unswervingly, whereas *mpendaji* is an occasional, irregular lover, someone who does not take his/her love seriously. Put differently, although both express the actor or the verber so to speak, they differ in terms of the level of commitment. *Mpenzi* is much more committed than *mpendaji*. This phenomenon is not idiosyncratic to the root *pend-* only, but is also characteristic of the majority of the verbs as the following set of data shows:

- (23) *jenga* 'build' *jenzi*¹¹ 'builder'
tunga 'compose' *tunzi* 'composer'
panda 'plant (seeds)' *panzi* 'planter of seeds'
unda 'construct' *unzi* 'constructor'
vua 'fish' *vuvi* 'fisherman'
lapa 'eat greedily' *lafi* 'greedy person'

On the other hand, the nominalizer *-e* is not used, instead, the verb root idiosyncratically uses the passive marker *-w-* to form the noun *mpendwa* 'the beloved one'. Another verbal root which behaves alike is *tuma* 'send'. This verb

root derives the noun *mtumwa* which means 'a slave'. Unlike *pend-* the verbal root *tum-* allows the derivation of *mtume* 'the sent one/prophet'. The verbal root *pend-* disallows *mpende* for the meaning 'the beloved one'. Yet, as a nominalizer *-e* is actively used by a number of verbal roots such as the following:

- | | |
|---|--|
| (24) <i>peta</i> 'bend' | <i>pete</i> 'ring/something bent' |
| <i>umba</i> 'create' | <i>kiumbe</i> 'something created/creature' |
| <i>shinda</i> 'win over someone/defeat' | <i>mshinde</i> 'someone defeated' |
| <i>nyonga</i> 'strangle/downtread' | <i>mnyonge</i> 'someone downtrodden' |
| <i>tuma</i> 'send' | <i>mtume</i> 'someone sent' |

Every verbal root in the language idiosyncratically chooses a number of nominalizers which participate in derivational processes involving it. The lexicographer cannot, by mere inspection of the forms, decide on the right nominalizers for the right verbs. Unfortunately the KKS does not point out which nominalizer is used by the root in each of its morphological processes.

What is underscored here is that verbal forms cannot be correctly interpreted without knowing their morphological history. Likewise, when suffix introduction rules are violated, semantic opacity results. When, for example, a stative suffix is introduced to stative verb roots or those which are intransitive, the resulting verbal forms are impossible to interpret semantically. The examples *chemkika* and *gumika* given above are good examples of semantic opacity. Let us now investigate verbal forms with more than one suffix.

4.2 More than one suffix

When a verbal root acquires a suffix, the resulting verbal form may be used as the input of another suffix introduction rule. Then, once it applies, this rule feeds another suffix introduction rule. In this way, some verbal forms consist of two or more extension suffixes. When the verbal forms under (3) above are allowed to acquire other suffixes, the resulting new forms are those under (25).

- (25) *pend-e-za* 'make someone love someone/something on account of'
pend-an-i-a 'make each other love for/at'
pend-e-an-a 'make each other love for/at'
pend-ek-e-a 'be lovable for/at'
pend-esh-an-a 'make each other lovable'

Except for the last, all these verbal forms make use of two suffixes. The first, *pendeza*, has the applicative followed by the causative suffix. The second, *pendania*, has the reciprocal followed by the applicative suffix. The applicative suffix follows the reciprocal in the third, *pendeana*. In the fourth, *pendekea*, the stative suffix follows the applicative. In the last verbal form, *pendeshana*, without going into the details of the morphophonemics alluded to earlier, the causative follows the reciprocal suffix. It will be recalled, however, that the introduction of the causative suffix sets in after the stative suffix has been introduced. Thus, to

begin with, we have *pendeka* which feeds the causative introduction rule to derive *pendesha* and then the causative is followed by a reciprocal suffix.

When two suffixes are used simultaneously, a number of morphosyntactic constraints tend to surface. One such constraint is that the applicative suffix cannot follow the stative. If they are used consecutively, the latter must precede the former. This claim is confirmed by the fact that the verbal form in (26)(a) below is unacceptable but (26)(b) is acceptable:

- (26) (a) **pend-e-ek-a* 'become loving for/on account of' (?)
 (b) *pend-ek-e-a* 'be/become lovable at/on account of'

Likewise, when the causative and the stative are used, the former precedes the latter. Both the applicative and the stative feed the causative introduction rule. The applicative suffix and the reciprocal suffix interchange positions without any consequences, both syntactic and semantic.

In addition to the morphosyntactic constraints, semantic problems surface as well. When two suffixes are simultaneously used, the semantic processing of the verbal form become somewhat enigmatic. Let us study the verbal forms under (25) in the sentences under (27):

- (27) (a) *Mtoto anapendeza* 'The child is attractive'
 (b) *Watoto wanapendania mali/kwao* 'The children like each other on account of wealth/home'
 (c) *Watoto wanapendeana mali* (as under (27)(b))
 (d) *Shati linapendekeke rangi yake* 'The shirt is lovable on account of its colour'
 (e) *Watoto wanapendeshana* 'The children are making each other lovable'

In order to process the above sentences accurately, we need to break them up into semantic kernels. In sentence (27)(a) *Mtoto anapendeza*, for example, the semantic kernels are:

- (i) Someone loves the child.
 (ii) There is a reason for loving the child.
 (iii) It is this reason that causes someone to love the child.

Thus, with these kernels the meaning of the sentence can be couched. Likewise the sentence (27)(b) *Watoto wanapendania mali/kwao* carries the following kernels:

- (i) The children love each other.
 (ii) Wealth is the reason for loving each other.
 (iii) Their home is where the loving of each other takes place.

What we need to note here is that the use of the applicative suffix introduces ambiguity into the sentences. Thus, *pendania* can mean 'love each other for' or 'love each other at'. The reading to be decided upon depends on what follows the verbal form. The former reading is arrived at if the verbal form is followed by *mali* 'wealth' and the latter is arrived at if *kwao* 'their home' follows.

Sentence (27)(c) *Watoto wanapendeana mali* is synonymous with (27)(b). The difference is more of a structural nature, because in (27)(b) the reciprocal precedes the applicative, whereas in (27)(c) the reverse is the case. When we say in Kiswahili *Shati linapendekea rangi yake*, sentence (27)(d), we mean to underscore three facts. First, someone somewhere loves a shirt. Second, the shirt has the specific feature making it lovable. Third, that specific feature is its colour. By using the two suffixes, namely the stative and the applicative, the speaker is able to effectuate the semantic content the sentence carries.

In (27)(e), the stative does not surface so that the causative *-sh-* precedes the reciprocal instead. The verbal form *pendeshana* expresses the stative, the causative and the reciprocal. Analyzed into its kernels, the sentence *Watoto wanapendeshana* means:

- (i) Each child is lovable to the other.
- (ii) Each child causes the other to be lovable.

When the two are combined, the verbal form expresses 'induced bidirectional lovability'. This in-depth understanding of the meaning of the verbal form puts the lexicographer in a better position to define it. Such a definition will be based on this bidirectional lovability. The definition of the verbal form in isolation can run somewhat as follows: *pendeshana* 'be involved in being lovable reciprocally/bidirectionally'.

Although definitions like these may be somewhat abstract, they prepare the ground for well-couched ones based on a thorough semantic analysis. In other words, verbal forms of this type cannot be cited in dictionaries without being adequately characterized.

The fact that verbal forms which combine more than one suffix are semantically intricate, constrains the number of affixes they acquire. On these grounds, each verbal root can only acquire a certain number of suffixes at a time. Theoretically, however, a root can acquire any number of suffixes. Some roots acquire as many as five suffixes as the root *pend-* does under (28):

(28) *pend-ek-e-z-e-an-a*

The semantic input of each suffix is as follows:

- ek-* 'be/become verbed'
- e-* 'for the purpose of/on account/for the sake of'
- z-* 'make/cause to become'
- e-* 'for the purpose of/on account/for the sake of'
- an-* 'each other'

The verbal form analyzed under (28) is a daily-used word in the language. In order to give a characterization, we need to adduce a principle in which each suffix has a specific scope. For example, the stative *-ek-* covers the root. Thus, *pendek-*, an extended root, expresses the state of being lovable, i.e. *pendek* means 'be lovable'. Then the *-e-* added to *pendek-* causes *pendekea* to express the mean-

ing 'be lovable for the sake of', etc. Let us now derive the meaning of the verbal form bit by bit:

- (29) *pend-* 'love'
pend-ek- 'be able to be loved/be lovable!'
pend-ek-e- 'be able to be loved for the sake of/on account of'
pend-ek-e-z- 'cause to be able to be loved for the sake of/on account of'
pend-ek-e-z-e- 'cause to be able to be loved for the sake of/at/on account of'
pend-ek-e-z-e-an- 'cause each other to be able to be loved for the sake of/at/on account of'

Note, however, that the applicative suffix is repeated. What we are somewhat uncertain about is whether or not, in each instance of its use, it expresses the same meaning. In other words, we are not certain whether or not the applicative suffix in *pendekea* is different from the one in *pendekezea*.¹² We are still uncertain whether a suffix used twice retains its semantic force even when it is used for a second time in the same verbal form. Could it be the case that the next use neutralizes the previous use through a lexicalization process? Although these issues need to be investigated, they are beyond the scope of this article.

If these verbal forms are allowed to nominalize — and theoretically this is possible — the verbal root *pend-* acquires many nominal derivatives. The following are nouns in use by Kiswahili speakers in their daily communication:

- (30) *pendekezeano*
pendekezano
pendekezeanio

From the discussion above, these derivatives may prove to be opaque to a Kiswahili learner. Yet, these and other derivatives need to be indicated and characterized in the dictionary. What is more, even for an ordinary dictionary, the lexicographer is strongly advised to find a way of helping his/her users who come across forms like these for the first time. One suggestion that can be given, is to have both the suffixes and the nominalizers characterized somewhere, if it is impossible to characterize every extension and its derivatives.

5. Conclusion

The verbal root *pend-* is representative of verbs of Bantu origin in Kiswahili. It has both actual and possible word forms it can generate. A dictionary which charts out all these word forms goes a long way to prove that the lexical paucity purported to be characteristic of the language is more imagined than real. A well-compiled Kiswahili dictionary, especially for learners, must show this state of affairs. Indeed, the majority, if not all, Kiswahili speakers, need a dictionary of this type for native speakers *per se* are, in truth, few.

In order to achieve this goal, the Kiswahili lexicographer needs to apply the 'enter-them-all' approach proposed by De Schryver and Prinsloo (2001). This approach is not only useful but also user-friendly for a dictionary intended for learners and language developers. It is an approach which, according to these writers, allows the inclusion of 'all the verbs and their derivations likely to be looked up'. When, however, costs must be taken into account, the lexicographer might opt for the 'frequency-based tail slots', or, better still, a combination of the two proposed approaches.

Endnotes

1. According to Khamis (1972), there are eleven suffixes, two of which are said to be lexical, because they do not change the syntactic patterning of the verbal forms that result.
2. Kiango (2000) does not seem to differentiate between derivation and inflection. According to him, inflection is a class maintaining derivation, whereas derivation *par excellence* is a class-changing derivation.
3. Kiswahili speakers as a whole are not native speakers of the same level of competence. For many, Kiswahili is a classroom phenomenon, and, for a few who live in urban areas, it is also a street as well as a domestic phenomenon. It is important, however, to add here that the mass-media use of the language has gone a long way in spreading it.
4. This suffix is, according to this article, a variant of the stative suffix. The difference between *bomoka* and *bomoleka* is that the former is derived from *bomoa* whereas the latter is derived from the verbal form *bomolea* which results from the acquisition of the applicative suffix.
5. *Dondosha* means 'cause something to drop' rather than 'cause something to be picked up'. With careful interpretation, we arrive at the meaning in which the verbal form calls for an action in the opposite direction.
6. A rule is said to feed another if its application triggers the structural context needed for it to apply.
7. 'Boil each other' makes sense in English, because the verb 'boil' can be used both transitively and intransitively.
8. The way we interpret *somesha* is 'make someone read' but actually the book is made readable to someone. *Kusomesha mtoto shule* is 'to make the school accessible to the child'. It is not, we contend, 'to make the child read'.
9. When verbal forms with two vowels extend to acquire either the stative or the applicative suffix, they do so with an introduction of a lateral /l/. This causes them to have two possible readings: *jalika* is acceptable if it is derived from *jalia* meaning 'be full for the sake of'.
10. If the sequencing of the extensions in the KKS is anything to go by, *poleka* is acceptable, because it is derived from *polea*.
11. After nominalization, the derived noun stem acquires a noun prefix. Thus, after it has been derived, *jenzi* acquires a noun prefix such as *m-* for 'the person who verbs' and *u-* which denotes the verbing skill or process.
12. The suffix, as has already been pointed out, has a homonymic behaviour. It can mean: 'for the benefit of', 'to the detriment of', 'at/on/in' or 'with/by'. That is why it is sometimes referred to as the prepositional suffix.

References

- Bokamba, E.G.** 1981. *Aspects of Bantu Syntax*. Preliminary edition. Urbana-Champaign: University of Illinois, Department of Linguistics.
- De Schryver, G.-M. and D.J. Prinsloo.** 2001. Towards a Sound Lemmatisation Strategy for the Bantu Verb through the Use of Frequency-based Tail Slots — with Special Reference to Cilubà, Sepedi and Kiswahili. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001: 216-242, 372.
- Givón, T.** 1971. Some Historical Changes in the Noun-class System of Bantu: Their Possible Causes and Wider Implications. Kim, C.-W. and H. Stahlke (Eds.). 1971. *Papers in African Linguistics*: 34-54. Edmonton/Champaign: Linguistics Research Inc.
- Givón, T.** 1971. Historical Syntax and Synchronic Morphology: An Archaeologist's Field Trip. *Papers from the Seventh Regional Meeting, Chicago Linguistic Society, April 16–18, 1971*: 394-415. Chicago: Chicago Linguistic Society.
- Gleason, H.A.** 1967. The Relation of Lexicon and Grammar. Householder, F.W. and S. Saporta (Eds.). 1967: 85-102.
- Householder, F.W. and S. Saporta (Eds.).** 1967. *Problems in Lexicography*. Bloomington: Indiana University/The Hague: Mouton and Co.
- Keach, C.B.** 1985. *The Syntax and Interpretation of the Relative Clause Construction in Kiswahili*. New York/London: Garland Publishing.
- Kenstowicz, M. and C. Kisseberth.** 1979. *Generative Phonology: Description and Theory*. New York/London/Toronto: Academic Press.
- Khamis, A.M.** 1972. *A Consideration of Grammatical Roles in Swahili*. Unpublished M.Phil. Thesis. York: University of York.
- Kiang, J.G.** 2000. *Bantu Lexicography: A Critical Survey of the Principles and Process of Constructing Dictionary Entries*. Tokyo: Institute for the Study of Language and Cultures of Asia and Africa (ILCAA).
- Mdee, J.S. and H.J.M. Mwansoko (Eds.).** 2001. *Makala ya kongamano la kimataifa Kiswahili 2000. Proceedings*. Dar Es Salaam: TUKI, Chuo Kikuu cha Dar Es Salaam.
- Prinsloo, D.J. and G.-M. de Schryver.** 2001. Taking Dictionaries for Bantu Languages into the New Millennium — with Special Reference to Kiswahili, Sepedi and Isizulu. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001: 188-215.
- Radford, A.** 1997. *Syntactic Theory and the Structure of English: A Minimalist Approach*. Cambridge: Cambridge University Press.
- Kamusi ya Kiswahili Sanifu.* 2004. Dar Es Salaam: Oxford University Press.
- Vitale, A.J.** 1981. *Swahili Syntax*. Dordrecht, Holland/Annaminson, U.S.A.: Foris Publications.
- Weinreich, U.** 1967. Lexicographic Definition in Descriptive Semantics. Householder, F.W. and S. Saporta (Eds.). 1967: 25-44.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

E-dictionaries and Phonolexicographic Needs of EFL Users*

Włodzimierz Sobkowiak, *School of English, Adam Mickiewicz University, Poznań, Poland (sobkow@amu.edu.pl)*

Abstract: The phonetic aspect of (EFL) dictionaries is among the most seriously underrated and underdeveloped in (meta)lexicography. Pertinent bibliography is scant and even the best learner dictionaries are found wanting on a number of counts. This contribution is both a summary of my thirteen-year-long research into (pedagogical) phonolexicography and a look ahead. The current state-of-the-art in phonolexicography is presented with particular attention paid to how the leading pedagogical EFL e-dictionaries relate to the actual and potential phonolexicographic needs of their users, both students and teachers.

The main themes are: (a) the specificity of phonolexicographic needs of (Polish) EFL users, (b) phonetic representation, both graphic and acoustic, in dictionaries and its problems, (c) phonetic access, i.e. querying the contents of the dictionary via the phonetic code, (d) didactic aspects of phonolexicographic information, i.e. its use in teaching and learning (EFL) pronunciation.

Keywords: E-DICTIONARIES, EFL, ENGLISH, PHONOLEXICOGRAPHY, PHONETIC ACCESS, PRONUNCIATION IN DICTIONARIES

Opsomming: E-woordeboeke en fonoleksikografiese behoeftes van EVT-gebruikers. Die fonetiese aspek van (EVT-)woordeboeke is van die ernstigste onderskatte en onderontwikkelde in die (meta)leksikografie. Tersaaklike bibliografiese besonderhede is skaars en selfs die beste aanleerderswoordeboeke blyk gebrekkig in 'n aantal opsigte te wees. Hierdie bydrae is sowel 'n opsomming van my dertien-jaar-lange navorsing oor (opvoedkundige) fonoleksikografie as 'n blik vorentoe. Die huidige stand van sake in fonoleksikografie word aangebied met besondere aandag gewy aan hoe die toonaangewende opvoedkundige EVT-e-woordeboeke verband hou met die werklike en potensiele fonoleksikografiese behoeftes van hul gebruikers, sowel leerders as onderwysers.

Die hoofonderwerpe is (a) die spesifieke aard van die fonoleksikografiese behoeftes van (Poolse) EVT-gebruikers, (b) die fonetiese weergawe, sowel grafies as akoesties, in woordeboeke en

* This article was presented in a more multimedia form as a keynote address 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. I am grateful to the AFRILEX Board for inviting me to deliver this presentation, which allowed me to focus my thinking on the user aspects of electronic phonolexicography. I am also grateful to Proff. Arleta Adamska-Salaciak and Robert Lew for helpful comments on the draft of this paper. All errors and imperfections remain mine only.

die probleme daarmee, (c) die fonetiese toegang, d.w.s. die raadpleging van die inhoud van die woordeboek deur middel van die fonetiese kode, en (d) die didaktiese aspekte van die fonoleksikologiese inligting, d.w.s. die gebruik daarvan in die onderrig en leer van (EVT-)uitspraak.

Sleutelwoorde: E-WOORDEBOEKE, ENGELS, EVT, FONOLEKSIKOGRAFIE, FONETIESE TOEGANG, UITSPRAAK IN WOORDEBOEKE

1. Phonolexicography

1.1 State of the art

In a series of publications spanning the last thirteen years (see <http://ifa.amu.edu.pl/~swlodek/public.htm>), I have repeatedly pointed out that phonolexicography is almost completely ignored, both linguistically and lexicographically. This has been independently noticed (but not dealt with scientifically) by other researchers, such as Landau (2001: 126), who approvingly summarized Hulbert's 1955 views in the following way: "Dictionaries are less satisfactory in pronunciation than in spelling, meaning, or etymology. The record of the spoken language is difficult to acquire, difficult to transcribe accurately and unambiguously, difficult to represent understandably in a dictionary transcription, and in most cases of less interest to the user than other kinds of information."

Also lexicographically aware phoneticians like Gimson (1973: 115) have had uncomplimentary views on how pronunciation is treated in dictionaries: "Today, when we no longer regard speech as a degraded form of writing, the pronunciation entry in dictionaries [...] should be accorded much greater importance." "Unfortunately, the theory underlying the pronunciation component in a dictionary is too frequently difficult to discern."

This, rather sad, state of the art in phonolexicography has not changed noticeably over the last two decades, which have seen an unprecedented growth of lexicography and metalexicography alike. It is enough to browse the abundant proceedings of the biannual Euralex congress, for example, to notice the almost complete absence of any deeper phonolexicographic reflection. It is almost as if there were a consensus among researchers that the few questions that arise at the interface between dictionaries and phonetics have long since been answered, e.g. by Gimson 1973, Abercrombie 1978 or Wells 1985. As a matter of fact, however, these 'classic' treatments of phonolexicography have, doubtless by design, raised more questions than answers. Some of these questions have again been approached in my two books on phonolexicography (Sobkowiak 1999 and 2006), which are now — to the extent that I can ascertain — the only such full-length monographs in existence. In the following section these issues will be considered briefly.

1.2 Main issues

1.2.1 The design of phonetic transcription

One of the most vigorously debated issues in phonolexicography (and beyond: EFL, for example) has been the choice of phonetic transcription to represent the pronunciation of headwords. It is paradoxical that this rather technical question should raise so much controversy. After all — phonolexicographically speaking — there are many other, apparently more substantial, problems than whether chronemes (:), tildes (~) and bars (-) should be allowed or not in dictionary transcriptions, and why. It is fair to note, of course, that these typographic decisions may reflect much deeper preoccupations with the structure and function of phonetic representation (for example, the classic one: whether vowel timbre in English is best treated qualitatively or quantitatively in such pairs as /i/ and /i:/). By and large, however, the graphical salience of typography has tended to dominate the scene to the disadvantage of other issues.

Even without serious study, it is easy to notice that there is little unanimity as to the choice of phonetic transcription for lexicography across the Atlantic, Americans sticking to their long tradition of respelling (Paikeday 1993), only occasionally modified by the introduction of schwa and a few other IPA borrowings. The plethora of EFL dictionaries published in European countries, on the other hand, tend to try all kinds of non-IPA experiments in transcriptional user-friendliness, which are bound to lead to widely differing results, depending on the L1 of the target users (see Sobkowiak 1999: 183-200 for a discussion of Polish (Polish English) transcriptions). There are few experimental studies of dictionary user reactions to different transcription systems (Fraser 1997), and much more research is needed, especially in the area of EFL pedagogical dictionaries, before solid applicational conclusions can be drawn. All in all, the design of an effective phonetic transcription for lexicography remains an open issue.

1.2.2 The scope of phonetic transcription

Which dictionary text should be transcribed: headwords, run-ons, phrases, definitions, examples, equivalents (see Bogaards 2005: 23)? This question has traditionally been answered to minimize the scope of phonetic transcription to head entries only, doubtless for reasons of typographic ease and economy of space. Both considerations must be revised now in the era of electronic dictionaries, leaving pedagogical issues on stage. For example, would learners benefit from having definitions in a monolingual dictionary fully phonetically transcribed? How about examples of usage? After all, "the phonological behaviour of words in context and its representation deserves equal attention" (Magay

1979: 103). Would it be useful to learners to be able to view phonetic transcription of either at different phonostylistic levels, say: 'careful speech' and 'fast speech'? While such solutions could now easily be implemented in electronic pedagogical dictionaries, I am not aware of any research in this direction. Indeed, the only book-length treatment of dictionary definition phonetics (of *Macmillan English Dictionary for Advanced Learners* (MEDAL)) which has ever appeared is Sobkowiak 2006. No empirical user research was attempted here, however. The results are, nevertheless, very interesting, showing, among others, that there are ways to make such definitions phonetically more user-friendly to EFL learners by a judicious choice of defining vocabularies, phonetic control over phraseology and sandhi, audio recordings or synthesis, and others. As before, and much despite the common metalexigraphic belief, the issue of the microstructural scope of phonetic transcription remains open.

1.2.3 The choice of default accent and phonostylistic level

These are two of those phonolexicographic issues which have received most attention. All of the 'classic' treatments mentioned above addressed them. There seems to be a fair amount of agreement to the effect that (a) the accent of choice is 'Received Pronunciation' (RP) for British English and 'General American' (GA) for American English, and (b) the citation form of a given word is chosen for phonetic representation. There are solid arguments behind these choices, which mostly reduce to the conviction that offering the EFL users more variety in either of the two sociophonetic dimensions would not only be counterproductive, but virtually confusing, not to mention the ubiquitous BC (before computers) concerns about precious printing space.

Notice, however, that the implicit assumption in the avoidance of socio-phonetic variation in EFL dictionaries has been that the proper target users of EFL pedagogical dictionaries are *learners* only. They are — so the argument goes — bewildered enough by other aspects of dictionary use to better spare them this additional worry of having to choose the right phonostylistic form for the occasion. Besides, research has shown that little dictionary use is conducted for purposes of speaking, so in this situation there is little to be gained from phonostylistic advice. These are strong arguments, but they seem to completely ignore the phonolexicographic needs of EFL *teachers* and very advanced learners, who may well need phonetic representation of headwords (a) in other accents than the canonical RP and GA, (b) on a variety of phonostylistic levels. The teacher's dictionary will be treated at some more length below.

1.2.4 Phonolexicographic consistency

Ensuring the consistency of phonetic representation across the entries, i.e. in

the macrostructure of the dictionary, has been one of the preoccupations of (pedagogical) lexicographers: "It goes without saying that, whatever method of indicating pronunciation is adopted, it should be consistently and correctly used in the body of the work" (Abercrombie 1978: 124). And yet, even in this electronic era, with dictionaries compiled with sophisticated software tools from the very inception, there are still too many inconsistencies to ascribe to pure chance. Many examples are listed in my phonolexicographic review of the first edition of MEDAL (Sobkowiak 2003).

Dictionary makers have problems with phonolexicographic consistency on many levels, from "conformity of representations with the established phonological rules of the language" (Sobkowiak 1997a: 98) to aligning headword transcription with its audio representation. Such errors could be spotted quite easily at the design and production stage, should phonolexicography attract more attention than it now does in lexicographic circles.

From the user perspective, the results of phonolexicographic (or, for that matter, any other) inconsistency may be confusing or positively misleading. While dictionaries are not prototypical resources used for the acquisition of foreign pronunciation, they do convey phonetic information in a supposedly orderly manner. In their craving for rules, learners have the right to expect that also the phonetic content of their dictionaries is rule-governed, and that by frequent use of the dictionary they may expect assistance in acquiring it, be it explicitly or implicitly. If some word-final nasals in the dictionary are transcribed syllabically (to take my favourite example; see Sobkowiak 1999: 94-112) and others — in analogous phonetic contexts — are not, phonolexicographic consistency is compromised and the process of acquiring the pertinent phonological rule of English interfered with. To take another example, in MEDAL1, I discovered twenty headwords starting with /ænti-/ (e.g. *antibiotic*, *anticlrical*, *anticyclone*, *antidepressant*, *antifreeze*, etc.) and seven with /æntɪ-/ (*antibody*, *anticlockwise*, *antidote*, *antigen*, *Antipodean*, *the Antipodes*, *antithetical*). If there is any linguistic motivation for this variation, it is certainly too subtle for learners to discern.

1.2.5 Phonetic access

Traditional paper-based dictionaries could not be accessed phonetically. Just about the only access path in semasiological dictionaries was through alphabetically ordered headwords. Sometimes indexes of different types were provided to aid users having less orthodox needs. It was quite recently that phonolexicographic needs of EFL users were finally recognized, and pedagogical e-dictionaries started to offer access to the lexicographic content through phonetic transcription, technically mediated by the so-called 'phonetic keyboards'. One such keyboard is reproduced below.

Figure 1: MEDAL1 *SoundSearch* phonetic keyboard

This is certainly a step in the right direction. Learners can now scan the dictionary for problematic homophones or phoneme clusters; teachers have an excellent tool for material and test preparation. As with many such new developments in contemporary lexicography, it remains to be seen just how user-friendly and useful these new access methods are in the actual EFL practice. Needless to say, there is practically no research yet in this area.

1.2.6 Audio sound representation

Another recent development of potentially significant impact on phonolexicography is good-quality Text-to-Speech Synthesis (TTS). Some systems now offer artificial speech virtually indistinguishable from natural. With the exception of some experimental applications, however, TTS has not yet been used in pedagogical lexicography. The potential benefits go far beyond economy of reader cost and space on the CD-ROMs, now taken up by the memory-hungry recordings of human speech. As I explained in Sobkowiak 1999, chapter 3.10, TTS could be used to audialize (a) any desired stretch of dictionary microstructure, apart from the headword itself, (b) at any desired phonolexical and phonostylistic level, (c) with the desired gender and voice quality, (d) with or without the desired L1 accent, and (e) for mere representation or as part of a suite of pronunciation exercises built into the dictionary. This phonolexicographic flexibility would be an enormous asset to both learners and teachers, who must currently resign themselves to recordings made by a handful of speakers, sometimes with their own phonetic idiosyncrasies. The one-word-

one-phonetic-representation principle of current lexicography is, of course, a serious misrepresentation of linguistic reality.

1.2.7 The typographic design of phonetic representation

Finally, another phonolexicographic issue which has received by far too little serious interest in the relevant literature is the screen typography of phonetic representation. This includes not only phonetic transcription proper, but also the accompanying symbols and flags, fonts with their attributes, scalability, GUI (Graphical User Interfaces) design, platform and copy-and-pasting robustness, and many others. Some examples of common problems in this area, with their consequences for the dictionary use by learners, are provided below (section 3.3). Needless to say, this is again a phonolexicographic field in urgent need of empirical user-directed research.

2. Phonolexicographic needs of EFL users

2.1 Are there any?

There is now abundant research to show that many EFL dictionary users have rather modest needs when it comes to phonetic representation (see Sobkowiak 1999: 115-121 and Lew 2004: 21-24 for overviews). Obviously, some of the largest groups of EFL dictionary users — translators, businessmen, secretaries, scientists — seldom need pronunciation in their lexical resources for many of their job-related activities. In heavy dictionary use for professional purposes, where the word's pronunciation is irrelevant and can actually be completely ignored, phonetic transcription can be worse than superfluous — it can be obtrusive. (Notice in this context that there is no option, in e.g. MEDAL1, to switch transcription off completely for screen display.)

Also some EFL learners express their disinterest in matters phonetic in dictionaries. This can well be understood in a school situation where reference resources are mostly used to practice the so-called 'passive' skills: listening and reading, with emphasis on the latter. There are clear logistic problems in using dictionaries for speech: listening and speaking alike. All this acknowledged, it is interesting that Lew (2004: 111-112) found as many as 28% (Polish) learners in his study claiming that they were consulting their dictionaries for pronunciation *often* and *most often*; see Table 1 below. Additionally, the phonetic use of the dictionary went up sharply in frequency at the most advanced proficiency level (Lew 2004: 123).

Table 1: Phonetic consultations of dictionaries in Lew 2004

never	rarely	often	most often	missing
179 (25%)	336 (47%)	160 (23%)	35 (5%)	2 (0%)

All these results show that there is a stable minority of EFL learners, many of them advanced, for whom phonetic representation in dictionaries does matter. As before, an important proviso needs to be mentioned: nobody has yet studied the phonolexicographic needs of EFL *teachers*. Lew's proficiency-stratified results suggest that these may be quite extensive.

My own research in phonolexicographic needs of Poznań university students (Sobkowiak 1999, part 2) points in a similar direction (although this sample may not have been quite representative of the population of advanced EFL learners because of the heavy phonetic bias of the School of English EFL programme). The following is a subset of questions asked in my questionnaire returned by more than six hundred students of English, at the Adam Mickiewicz University, Poznań, Poland.

Table 2: Some phonolexicographic needs of EFL Poznań university students (Sobkowiak 1999)

Statement	yes	no	?
English dictionaries should show pronunciation in some way	601	24	20
I check pronunciation of some words when I read	547	95	3
I have recently checked pronunciation in an English dictionary	611	32	2
In listening comprehension tasks, I look up some words I hear	458	170	17
International Phonetic Transcription (IPA) should be used in all English dictionaries	499	61	85
Learners' dictionaries should address their common pronunciation problems	528	51	66
Separate pronunciation dictionaries are nonsense	90	519	36
In a multimedia computer dictionary, phonetic transcription is useless	33	372	240

2.2 Learner-friendliness

With the advent of computers, especially after GUIs had been introduced, there has been a lot of preoccupation with user-friendliness ("the elegance and clarity with which the interaction with a computer program or a web site is designed" — Wikipedia). In the present context, the issue is (EFL) learner-friendliness, of course, but the design criteria remain the same.

There are many different respects in which dictionaries can be phonolexicographically learner-friendly. If lexicographers are aiming at global 'user-friendliness' of their defining vocabulary, they should certainly also make it phonetically friendly by avoiding notoriously phonetically difficult words, for example. As I showed in Sobkowiak 2006, there are few signs that this aspect of defining vocabularies comes under deliberate lexicographic control in the leading EFL dictionaries now on the market. And yet, phonetically difficult words in dictionary definitions will tend to impede the reading and understanding

process, particularly in those learners who continue to vocalise or articulate subvocally in silent reading, as many do (cf. e.g. Gibson and Levin 1975: 342, Grodziński 1976: 52ff).

EFL electronic dictionaries could dynamically adjust their definitions to the learner's needs and requirements, not only in terms of their lexical scope (defining vocabulary) and syntax, but also in terms of pronunciation (see De Schryver 2003 for this and other lexicographers' dreams). For example, if *thorough* is among the phonetically hardest lexical items for EFL learners, why not use a substitute in definitions (e.g. *complete*) adjusted for pre-intermediate learners, where (syntactically, stylistically and pragmatically) appropriate? Or at least why not reduce the definition incidence of *thorough* (it now stands at 22 in MEDAL1)?

2.3 Teacher's dictionary?

Google search (in website title) for the phrase "teacher's dictionary" currently (7 October 2006) yields 11 hits, none of which is actually a pedagogical EFL dictionary for teachers in the sense which has been of concern here. "Learner's dictionary", on the other hand, yields 21 900 web pages. Why this difference? Do EFL teachers not use EFL dictionaries? Of course, they do: they use their learners' dictionaries, which are in many respects unfit for teachers' needs. What would, then, be the expected features of an EFL teacher's e-dictionary, from the phonolexicographic perspective? For lack of space (and the issue quite clearly deserves an in-depth treatment), they will briefly be listed below. A teacher's dictionary must offer:

- liberal deployment of phonetic transcription, both for representation and indexing,
- powerful multicriterial phonetic access mechanisms,
- phonetically streamlined word-list generation functionalities,
- robust copy-and-paste facilities, and
- indication of accentual, phonostylistic and phonolectal variation, with
- advice on the preferred phonetic form in various situations of use,
- guidance on likely phonetic problems with the given entry/definition/example, and
- remedial drills and exercises, generated automatically from phonolexicographic context.

This is certainly a far-from-exhaustive list of desiderata. The lexicographic theory (let alone practice) of the teacher's dictionary does not exist. I believe that, with half the globe learning and *teaching* English as a lingua franca, it is now long overdue.

3. Phonetic representation

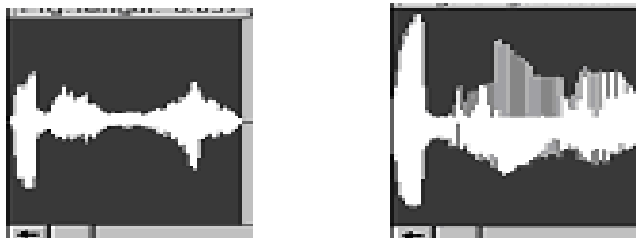
3.1 Transcription vs. others

Doubts have sometimes been voiced as to the desirability of phonetic transcription in dictionaries for learners (Kernerman 1996: 409), in view of the now pervasive use of recorded speech as phonetic representation of headwords. Again, little pertinent research has been conducted in this area. It is intuitively obvious that learners are indeed quite fond of clicking on the headword to listen to the model pronunciation. It is less obvious how many of them would want to keep the transcription as well. The answer, which I obtained in my 1999 questionnaire, listed in Table 2 above ("In a multimedia computer dictionary, phonetic transcription is useless": yes – 33, no – 372, don't know – 240), is hardly indicative of the needs of 'ordinary' EFL learners, i.e. ones of non-academic proficiency, who do not study to become teachers of English. Notice also, that while most subjects claimed that even in a multimedia dictionary phonetic transcription is useful, quite a few (37%) were uncertain.

And yet, there are pedagogically clear advantages of phonetic transcription displayed in addition to the audio representation. While the latter provides a receptive and productive model to be used for immediate mimicry, the former offers categorization, systematization and 'accuracy anchor', to borrow Grabe and Stoller's (1997: 112) term.

Similarly, a visual waveform representation of spoken words, increasingly more often provided as a thrilling multimedia extra, may offer multisensory feedback for practice, but is risky because "the *visual* comparison of the two sound waves, the model's and the learner's repetition [...] is at best inconsequential, and at worst thoroughly misleading and frustrating. The graphic representation of a waveform has a very complex relationship to its acoustic basis, and the latter to both its articulation and perception" (Sobkowiak 1997b: 335). Consider, for example, my own best rendition of *Good morning* compared with the native model in Figure 2. Should I not have known better, I would have been very frustrated by the dissimilarities between the two waveforms, apparently testifying to my complete mispronunciation of this phonetically easy phrase.

Figure 2: A waveform of *Good morning*; model left, learner right



3.2 Simplified transcription?

To foreign learners there are no self-pronouncing words. For a quarter of a century, or more, my own phonetic image of the word *desultory* has been parallel to *compulsory*, i.e. */də'sʌltəri/. *Instead* may look self-pronouncing to native speakers of English, but is far from this, as many beginning/intermediate Polish errors testify, e.g. */m'sti:d/. All headwords in an EFL pedagogical dictionary should have phonetic representation, however obvious it may seem to the (native) dictionary makers.

As mentioned in section 1.2.1, while phonetic transcription has definitely been normalizing during the whole 20th century, there is still wide diversity, especially cross-Atlantically. Respelling continues to be quite popular in America, where the concept of an EFL monolingual learner's dictionary with the 'international' IPA transcription is a relatively recent graft from Great Britain. From the point of an EFL dictionary user, however, respelling can be thoroughly confusing, as it is of course heavily target-language-dependent. The following praise, voiced on <http://dictionary.reference.com/help/faq/language/s36.html>, can then only refer to native speakers of English: "A respelling pronunciation system is fairly practical. No special characters or diacritics are used. No pronunciation guide must be relied upon. Examples are: *accident* (AK-sidunht), *diamond* (DIE-muhnd), *garage* (gah-RAHZH, guh-RAHZH, GA-rahzh)." A Polish EFL learner would be uncertain as to the correct phonetic value of /AK/, /dunht/ or /RAHZH/. The bottom line is, then, that in the EFL situation any such 'simplifications' are very risky and should be thoroughly tested empirically before deployment. It is not surprising that the IPA standard, despite much criticism directed at it from EFL circles (especially in the 1970s) holds on quite staunchly in Europe.

An important proviso in this context is that in e-dictionaries many types of transcription are in principle mutually convertible and selectable. There is thus nothing to stop EFL e-dictionary makers from offering learners a transcription of their choice, including the many subtle varieties of IPA on the one hand, and some L1-sensitive derivative simplifications on the other (see, e.g. Sobkowiak 1997c).

3.3 Typography

As mentioned briefly in 1.2.7, the whole issue of typographic design of phonetic transcription for screen rendering has so far received very little attention. For example: "the whole of CD-ROM MEDAL is set in sans serif Arial-like font [...]; the only exception being the non-Roman IPA symbols [...] which are serifed. Depending on the graphics card's selected screen resolution, both the shape and the size of the transcription field symbols may differ: /'kɒntrəvɜːsi/" (Sobkowiak 2003: 427). This is indeed a common weakness of e-dictionaries, including the bilingual English–Polish dictionaries which I have seen. The con-

sequences for the user/learner are twofold: (a) font inconsistencies show that phonetic representation is somehow inherently sloppier than the rest of the microstructure, hence (maybe) of less lexicographic importance; and (b) subtle differences in symbol shape may be taken to imply phonetic contrasts (after all, in transcription, unlike in orthography, every little bit counts).

Another weakness of e-dictionaries is that transcription does not survive copy-and-pasting, a common enough procedure, especially with young computer users, who usually type better than they handwrite from kindergarten onwards. The following is the phonetic fragment of the *controversy* entry copied from MEDAL1 on CD-ROM into MS Word 2000: / B `k\197ntrEv3:si, B kEn`tr\197vEsi /. While the mapping with IPA may be one-to-one, it would certainly be too much to expect learners to translate from this SAMPA-like transcription to anything closer to standard.

3.4 Choice of model

As mentioned above, the traditional monolectal solutions to the problem of pronunciation model choice appear to work best for most, especially lower-proficiency, learners. On the other hand, real phonetic variation caused by dialect, phonic style, lexical frequency, or indeed the so-called 'free variation' should not be suppressed by the overwhelming force of the "monostylistic curse" (Bailey 1986: 26). Advanced learners and teachers may need this information to follow their aims. The solution appears to be to customize display of phonetic content to the needs of the user. Just as different transcription types can be switchable in an e-dictionary, so can different phonolects. To the extent that they are phonologically rule-governed, the switching can be done automatically across the whole macrostructure of the dictionary. Idiosyncrasies would need to be programmed individually for each word, i.e. the way that all dictionary transcription is prepared nowadays.

3.5 Transcription vs. audio

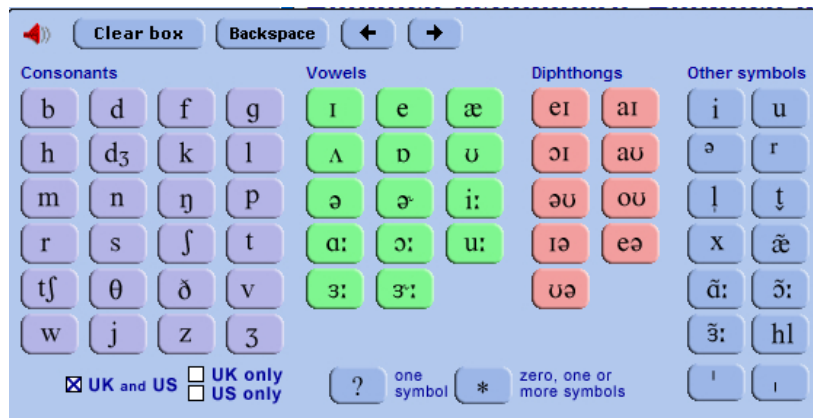
The main problems with the recorded audio component of many dictionaries are: (a) the list-reading effects (e.g. contrastive stress, sustained intonation contours), (b) the occasional mismatch between recording and transcription, (c) limiting the recordings to headwords only (the recent edition of *Longman Dictionary of Contemporary English* is an exception here: http://www.longman.com/ldoce/about_cd.html). As an example of (b), listening to all British headword recordings in MEDAL1 within the ⟨a⟩-letter range, I found about seventy cases of mismatch in the few thousand entries, e.g. *academician* is transcribed as /əˌkædəˈmɪʃn/ but pronounced as /æˌkædəˈmɪʃn/ in MEDAL1. Or consider the non-coalesced recording of *aperture* /'æpətʃə/. Good quality TTS, as mentioned above in section 1.2.6, would solve all of these problems.

4. Phonetic access

In the traditional paper dictionaries, transcription fulfils but one function — representational. Learners refer to it to find out about the pronunciation of the (orthographically) located word. This is very different from the indexical (query) function, which transcription additionally has in a phonetic-access enabled e-dictionary, such as MEDAL. Here learners enter phonetic transcription for a word or word-list which they search. This function of transcription simply did not exist in traditional dictionaries. In e-dictionaries, both functions must be smoothly integrated. As it appears, this is far from trivial. Consider the question of grapho-phonemic biuniqueness for data entry, for example. From the point of view of mere representation, it does not matter whether English affricates or diphthongs are coded mono- or bisegmentally. In their representational function /tʃ/ and /tʃ/, they do not differ at all. Bisegmental coding for querying (indexical function) will not do, however. The search for /*ʃ/ words will yield: *aitch, approach, arch, attach*, etc., as it does in MEDAL1, for instance. As I described in Sobkowiak 2003, there are many other such problems, concerning stress marks, wildcards and brackets, for example.

Some EFL-wise useful queries, such as one for aspirated plosives, are quite simply impossible, not only in the electronic versions of standard learners' dictionaries, such as MEDAL or Longman, but even in the dedicated *pronouncing* dictionary, such as the *Cambridge English Pronouncing Dictionary* (CEPD) on CD-ROM (see my review Sobkowiak 2005). As can be seen in the screenshot of the CEPD phonetic keyboard in Figure 3, such EFL-wise exoticisms as the Welsh voiceless laterals can be entered, but there is no way to type aspiration or sonorant syllabicity, both phenomena of obvious interest to EFL learners. The conclusion is that phonetic access has yet some way to go before it can be fully and profitably integrated in EFL learners' e-dictionaries.

Figure 3: CEPD phonetic keyboard



5. Didactic aspects of phonolexicography

5.1 Dictionaries as teaching/learning resources?

Despite some voices to the contrary: "It cannot realistically be seen as part of the dictionary's function to teach the sound system" (Brazil 1987: 161), EFL dictionaries are gradually transformed into one-stop learning resources, with grammatical, pragmatic and phonetic advice on board, a variety of self-study facilities, multiple suites of exercises and hyperlinks to even more support on the world-wide web. In Sobkowiak 1998, I argued that despite the traditional idiographic perspective of lexicography, EFL MRDs (machine-readable dictionaries) can indeed teach foreign pronunciation. This is all the more true of e-dictionaries, with their phonolexicographic query potential, which I have been unveiling in this article. For example, which EFL resource other than properly coded e-dictionary could (semi-)automatically generate exercises spun around: (a) those English animal names which are (b) relatively common in colloquial English, but (c) relatively difficult (grapho)phonetically to Polish (foreign?) learners: *calf, lamb, sow, bison, donkey, giraffe, leopard, monkey, reindeer?*

5.2 Techniques

The only type of pronunciation exercise currently used in EFL e-dictionaries is "repeat after me", as in the following screenshot from MEDAL1.

Figure 4: "Repeat after me" exercise in MEDAL1



This limitation, however, is not one of technology or pedagogy, but of imagination. For example, pronunciation practice could easily be combined with flashcards. Why is it that the only elements of an entry's microstructure used in flashcards are its headword and definition? Why not let the learner "type the mystery word" in response to its phonetic transcription or recording (dictation), or both, as well as to its definition? Why not flash the headword and ask the learner "Do you remember how to pronounce this word?", as well as "Do you remember what this word means?", the only option now built in?

But one could, of course, go much further. The wildcard and regular-expression queries of phonolexicographic content yield lists of items from which the original criterion could be inductively reconstructed by the learner. For the beginner: "What do these words have in common, as far as pronunciation goes: *dough, go, know, sew, toe?*" (a selection from forty-nine MEDAL1 /?əʊ/

headwords). For the advanced: *aphrodisiac, Chianti, cordiality, piano, react?*" (a selection from ninety-nine /*iæ*/ headwords, containing a particularly recalcitrant case of vocalic hiatus).

Of course, it is easy to transform these exercises into binary- or multiple-choice format, matching, selection or minimal pairs. It is possible to combine them with audio and/or transcription, as well as part-of-speech information, lexical frequency tagging, dialectal and stylistic stratification, illustrations, etc.

5.3 L1-sensitivity

Both theoretical and applied phonology must be L1-sensitive because phonetic interference from L1 is the amplest source of L2 pronunciation errors on most levels of proficiency. Contemporary EFL dictionaries are not L1-sensitive, or only superficially so (e.g. false-friends lists), mostly due to the overwhelming commercial factors. For example: the EFL dictionaries currently available on the Polish market do not take account of one interference-prone Polish graphophonemic rule: ⟨c⟩ → /ts/. To Polish beginners, *romantic* is /ro'mantits/. There is an abundant bibliography of sources on common Polish phonetic errors at different proficiency levels. All that needs to be done is to build this knowledge into EFL e-dictionaries for Polish learners, and then use it to post cautionary flags or construct phonetically L1-sensitive exercises using the phonolexicographic content of the dictionary as input. This is already happening on other levels of lexical structure in dictionaries: morphological, syntactic, pragmatic. Phonetics is obviously lagging behind.

5.4 The undiscovered potential of definitions

In the final chapter of Sobkowiak 2006, I unveil some of the so-far undiscovered pedagogical potential of monolingual EFL learners' dictionary definitions. If treated as a properly phonolexicographically annotated corpus of text, rather than separate text bits attached to headwords, they can furnish an excellently rich resource for learners, providing opportunity for pronunciation practice along different phonetic dimensions: from segmentals, through sandhi, to stress, rhythm and intonation.

Consider an example. A learner who has problems with linking-r, dental(ized) clusters or fast-speech /d+j/ affrication (and many invariably do), such as in this MEDAL1 definition of *melt* – *if you melt into or against someone you relax as they hold you close in a romantic way*, could obtain some advice and ample material for practice taken from other phonetically treated definitions, where such phonetically troublesome phenomena would be suitably coded and could be retrieved in the form of a phonetic concordance focused on the relevant difficulties, such as in the following example:

- exactly: *in every way or every detail*
- intense: *very great or extreme*

- severe/ly: *very strict or extr~~e~~me*
- to have one foot in the grave: *to be very old or ill and likely to die soon*

6. More (phono)lexicographic dreams

6.1 Phonetically treated definitions

I dream of dictionaries which would redress the anti-phonetic bias of current lexicography. In such dictionaries, not only the phonetic representation of the headword would be carefully thought over, but the entire entry would likely receive phonolexicographic attention. Definitions would be designed and written according to some phonetic guidelines, just as they are written according to strict syntactic and stylistic guidelines today. In consequence, such definitions would be easier to read, both as meta-text and text, and the incidental learning of vocabulary would get a boost. Properly annotated, they could be used as a (phonetic) electronic corpus resource in simple lookup as well as in a variety of sophisticated queries informing word-list generation, test preparation, materials design, etc. (see 5.4 above). Audio-recorded or speech-synthesized, they could make an e-dictionary a yet more functional didactic aid, with all the multimedia involvement expected by contemporary learners (Sobkowiak 2006).

6.2 Articulatory animation

The so-far lexicographically unimplemented type of potentially pedagogically useful phonetic representation is articulatory animation. Realistic avatars of the *Baldi* kind (<http://mambo.ucsc.edu/pdf/WilliamsRDD.pdf>) could be used, with speech animated in real time. Different perspectives, zooms, transparencies and tempos could be used (see Figure 5 below). The animation can be coordinated with a human recording or with TTS output. The technology is now mature for use in e-dictionaries. The benefits for e-dictionary EFL users are too obvious to elaborate on.

Figure 5: *Baldi* going transparent



7. Instead of conclusions

As was alluded to above, "indicating pronunciation is often under-estimated by the critics of dictionaries as being a derivative business" (Magay 1979: 99). I hope to have shown that phonolexicography, while far from being "a derivative business", can and should become a thriving scholarly field, as well as a fully-fledged discipline of applied lexicography. Many users use their EFL dictionaries for pronunciation, starting from the modest one-off lookup of the word's phonetic transcription, through phonetic-access wildcard searches for troublesome clusters, to pronunciation exercises built into some e-dictionaries for learners. Despite its enormous pedagogical potential, and unlike other sub-disciplines of learners' lexicography, theoretical and applied phonolexicography has hardly been developing in the last decades. This is one more attempt to start to change this unsatisfactory situation.

Bibliography

- Abercrombie, D.** 1978. The Indication of Pronunciation in Reference Books. Stevens, P. (Ed.). 1978. *In Honour of A.S.Hornby*: 119-126. Oxford: Oxford University Press.
- Bailey, C.-J.N.** 1986. Even Dialectologists Get the Blues: Inadequate Phonetic Transcription of English. *Papere zur Linguistik* 35(2): 3-38.
- Bogaards, P.** 2005. Dictionaries and Productive Tasks in a Foreign Language. *Kernerman Dictionary News* 13: 20-23.
- Brazil, D.** 1987. Representing Pronunciation. 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*: 160-166. London/Glasgow: Collins ELT.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Fraser, H.** 1997. Dictionary Pronunciation Guides for English. *International Journal of Lexicography* 10(3): 181-208.
- Gibson, E.J. and H. Levin.** 1975. *The Psychology of Reading*. Cambridge, Mass.: The MIT Press.
- Gimson, A.C.** 1973. Phonology and the Lexicographer. McDavid, R.I. Jr and A.R. Duckert (Eds.). 1973. *Lexicography in English*: 115-121. International Colloquium, New York 1992. *Annals of the New York Academy of Sciences* 211. New York: Academy of Sciences.
- Grabe, W. and F.L. Stoller.** 1997. Reading and Vocabulary Development in a Second Language. Coady, J. and T. Huckin (Eds.). 1997. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*: 98-122. Cambridge: Cambridge University Press.
- Grodziński, E.** 1976. *Mowa wewnętrzna* [Inner speech]. Wrocław: Ossolineum.
- Hulbert, J.R.** 1955. *Dictionaries: British and American*. London: Deutsch.
- Kernerman, L.** 1996. English Learners' Dictionaries: How Much Do We Know about Their Use? Gellerstam, M. et al. (Eds.). 1996. *Euralex '96 Proceedings*: 405-414. Göteborg: Göteborg University, Department of Swedish.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.

- Lew, R.** 2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań: Motivex.
- Magay, T.** 1979. Problems of Indicating Pronunciation in Bilingual Dictionaries with English as the Source Language. Hartmann, R.R.K. (Ed.). 1979. *Dictionaries and Their Users*: 98-103. Proceedings of the 1978 BAAL Seminar on Lexicography. Exeter Linguistic Studies 4. Exeter: Exeter University Press.
- Paikeday, T.M.** 1993. Who Needs IPA? *English Today* 9(1): 38-42.
- Sobkowiak, W.** 1997a. Consistency in EFL Dictionary Phonetics. Waniek-Klimczak, E. (Ed.). 1997. *Teaching English Phonetics and Phonology II. Accents '97*: 95-102. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Sobkowiak, W.** 1997b. Speech in EFL CALL. Strykowski, W. (Ed.). 1997. *Media a edukacja*: 329-339. Poznań: eMPI².
- Sobkowiak, W.** 1997c. Radically Simplified Phonetic Transcription for Polish Speakers. Hickey, R. and S. Puppel (Eds.). 1997. *Language History and Linguistic Modelling. Festschrift for Jacek Fisiak on his 60th Birthday: 1801-1830*. Berlin: Mouton.
- Sobkowiak, W.** 1998. Can EFL MRDs Teach Pronunciation? Fontenelle, T. et al. (Eds.). 1998. *Euralex '98 Proceedings*: 271-277. Liège: University of Liège, English and Dutch Departments.
- Sobkowiak, W.** 1999. *Pronunciation in EFL Machine-Readable Dictionaries*. Poznań: Motivex.
- Sobkowiak, W.** 2003. Pronunciation in *Macmillan English Dictionary for Advanced Learners* on CD-ROM. *International Journal of Lexicography* 16(4): 423-441.
- Sobkowiak, W.** 2005. Review of *Cambridge English Pronouncing Dictionary* (CEPD) on CD-ROM. *TESL-EJ* 8.4. <http://writing.berkeley.edu/TESL-EJ/ej32/m1.html>.
- Sobkowiak, W.** 2006. *Phonetics of EFL Dictionary Definitions*. Poznań: Wydawnictwo Poznańskie.
- Wells, J.C.** 1985. English Pronunciation and its Dictionary Representation. Ilson, R. (Ed.). 1985. *Dictionaries, Lexicography and Language Learning*: 45-51. London: Pergamon.

Issues in the Planning and Design of a Bilingual (English–Northern Sotho) Explanatory Dictionary for Industrial Electronics

Elsabé Taljard, *Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (elsabe.taljard@up.ac.za),*

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

and

Liam A. Gauton, *NamITech Pty. Ltd. and Department of African Languages, University of Pretoria, Pretoria, Republic of South Africa (liam.gauton@namitech.com)*

Abstract: The focus of this article is the planning and design of a bilingual, explanatory dictionary for industrial electronics with a clearly delimited and very specific target user in mind. Since the number of lemmas to be treated in the dictionary is limited to 500, special care must be taken to select those lemmas that are relevant for both the purpose of the dictionary and the needs of the target user. It is indicated that the main consideration in the planning of the envisaged dictionary is user-friendliness, as dictated by the intended target users. In this article, a novel approach to the semi-automatic selection of lemmas for inclusion in an LSP dictionary is described. The procedure that is used for the extraction of definitional information from an electronic corpus is also explained.

Keywords: LSP LEXICOGRAPHY, DICTIONARY PLANNING, LEMMA SELECTION, SEMI-AUTOMATIC TERM EXTRACTION, DEFINITIONAL INFORMATION, INDUSTRIAL ELECTRONICS, CORPUS-BASED LEXICOGRAPHY

Opsomming: Kwessies by die beplanning en ontwerp van 'n tweetalige (Engels–Noord-Sotho) verklarende woordeboek vir industriële elektronika.

Die fokus van hierdie artikel is die beplanning en ontwerp van 'n tweetalige, verklarende woordeboek vir industriële elektronika met 'n duidelik afgebakende en baie spesifieke teikengebruiker in gedagte. Aangesien die getal lemmas vir behandeling in die woordeboek tot 500 beperk is, moet besondere sorg gedra word dat daardie lemmas gekies word wat beantwoord aan die doel van die woordeboek én die behoeftes van die teikengebruiker. Daar word uitgewys dat die hoofoorweging in die beplanning van die beoogde woordeboek gebruikersvriendelikheid is, soos bepaal deur die bestemde teikengebruikers. In hierdie artikel word 'n nuwe benadering tot die semi-outomatiese keuse van lemmas vir insluiting in 'n vakwoordeboek beskryf. Die prosedure wat vir die onttrekking van definisiële inligting uit 'n elektroniese korpus gebruik word, word ook verduidelik.

Sleutelwoorde: LEKSIKOGRAFIE VIR VAKWOORDEBOEKE, WOORDEBOEKBEPLANNING, LEMMASELEKSIE, SEMI-OUTOMATIESE TERMONTTREKKING, DEFINISIËLE INLIGTING, INDUSTRIËLE ELEKTRONIKA, KORPUS-GEBASEERDE LEKSIKOGRAFIE

1. Introduction

The research upon which this article is based, forms part of a larger project entitled Language, Educational Effectiveness and Economic Outcomes, thus LE³O, done under the auspices of the Centre for Research in the Politics of Language (CentRePoL) at the University of Pretoria. The main aim of this project as stated by Webb (2005: 5) is to contribute towards addressing poverty and inequality in the distribution of wealth through the development of the vocational literacy of workers. This is done by "researching the role of language in vocational education and training, proposing ways in which the fundamental role of language in vocational training and practice can be recognised and corrected, and proposing specific policies regarding the role of language in Further Education and Training in South Africa". One of the subprojects is the introduction of Northern Sotho as medium of instruction in a course in industrial electronics (IE) taught at two colleges for Further Education and Training (FET) in the Greater Metropolitan Area of Tshwane. An offshoot of this project is the compilation of a bilingual (English–Northern Sotho) explanatory dictionary for industrial electronics (BEDIE). Due to financial and other constraints the number of terms to be treated in the dictionary is restricted to approximately 500. Due to this rather severe restriction, particular attention must be paid to the proper lexicographic and terminological planning of this (rather rudimentary) dictionary so that the maximum amount of relevant information can be imparted to the user in the most effective way. The planning of this dictionary therefore forms the focus of this article, with specific reference to the selection of the lemma list and the semi-automatic extraction of definitional information from an electronic English IE corpus. The actual compilation of the dictionary will take place in two phases. The first phase will consist of the selection of the English terms to be included in the dictionary, followed by the writing of terminological definitions for the selected terms. During the second phase, the terms and their definitions will be translated into Northern Sotho. The translation of the selected terms and their definitions from the original English source material into Northern Sotho is an undertaking generating its own pitfalls and challenges and does not fall within the ambit of the current investigation. (Cf. Gauton et al. (forthcoming) for an exposition of the difficulties inherent in this type of translation).

2. Intended purpose and function of the dictionary

Gouws and Prinsloo (2005: 13) point out that the purpose of a dictionary is determined by, *inter alia*, its typological nature and its intended target user

group. Typologically speaking, the envisaged dictionary will be a fully bilingual (English, Northern Sotho), unidirectional (English → Northern Sotho) explanatory dictionary. Taking the specific circumstances of the target user group into account, the dictionary will play an important role as a learning aid. The intention is to include the dictionary as an addendum of the textbook used in the instruction of industrial electronics so as to allow immediate access to the information provided in the dictionary. This would offer the users the opportunity of accessing definitions of basic industrial electronics terms in their home language, and would also provide them with Northern Sotho equivalents of key terms. The BEDIE is therefore primarily intended to assist learners with the decoding of texts written in a language foreign to them, i.e. English.

3. Intended target users

According to Bergenholtz and Tarp (1995: 20, 21), three crucial aspects need to be taken into account when the profile of the intended target user is drawn up, i.e. native-language competence, foreign-language competence and encyclopaedic knowledge. The intended target users of the BEDIE are learners aged 15–21 years, who are mother-tongue speakers of Northern Sotho, but who have hitherto received their formal education primarily through the medium of English, which is for most of them a second or even third language. Due to the fact that Northern Sotho has to a large extent not been used as a medium of instruction during their formal schooling, it is assumed that their academic proficiency in Northern Sotho is rather below the expected level, particularly within a technical field such as industrial electronics. This is confirmed by the preliminary results of an attitude test carried out by the LE³O team at the two colleges involved in this project, as reported by Webb (2005: 121): 41.7% of learners who indicated that their home language is Northern Sotho, responded by saying that they cannot speak Northern Sotho well; 60.8% indicated that they cannot read Northern Sotho well, whereas 70% indicated that they are not adequately proficient in writing Northern Sotho. It can therefore be concluded that the native-language competence of the intended target user is below standard. With regard to their proficiency in English, i.e. foreign-language competence, proficiency tests revealed that their proficiency in English is also way below par, as is evident from examples extracted from written work done by the students during this test. Compare a few randomly selected examples of written material produced by learners during the English Language Proficiency Test as cited in Webb (2005: 12) and shown in example (1).

- (1) — When we approaches the steps ... there was so many people try to get their class ... rember it was a serious injured ... school help by paying everything to him ... there is no job where you cold find one races only.

- I hope my situation will be requested soon ... my friend miss to step on the stairs and it was hurt badly.
- He doesn't help the customer well. When you give the bus drive the bus you must already teach him the ruel of the road and keep the customer good. And they must obey the ruel of the customer; I will happy with the manager her us.

It is interesting to note the discrepancy between learners' perception of their own competence and their actual proficiency in English — according to Webb (2005: 119), 92% of respondents reported that they spoke English well/very well, 96.8% indicated that they could read it well/very well, 90.7% were of the opinion that they could write English well/very well, whereas 99.1% indicated that their comprehension of English is good/very good. This is in stark contrast to the actual English usage, as reflected in the examples above.

With regard to their encyclopaedic knowledge, it must be kept in mind that learners entering an FET college have had little, if any, exposure to a subject such as industrial electronics. Secondly, taking available assessment results into consideration, it must be concluded that even after training the intended target user has an average to low level of encyclopaedic knowledge and can therefore be regarded as a layperson. Compare in this regard Table 1 showing the results obtained for industrial electronics during the first three years of training at the two colleges, as cited by Webb (2005: 50).

Table 1: Assessment results, trimester 2, 2002

	Average %			Total average %
	N1 (Grade 10)	N2 (Grade 11)	N3 (Grade 12)	
College A	43	43	43	43
College B	75	65	44	61

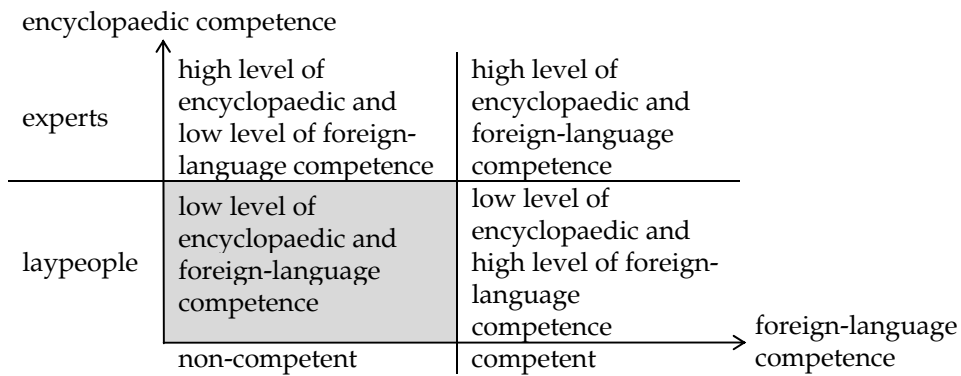
In terms of the four main types in user profile as distinguished by Bergenholtz and Tarp (1995: 21), the user profile for this particular project is indicated in Figure 1.

As Bergenholtz and Tarp (1995: 21) point out, it needs to be kept in mind that the transition between the different types is fluid.

Taking the broader South African situation with regard to the culture of dictionary use into account, it can furthermore be assumed that the target users have had very little, if any, exposure to the culture of dictionary use. According to Gouws and Prinsloo (2005: 42), the majority of South Africans find themselves in a "pre-dictionary culture environment". This is especially valid for speakers of the South African Bantu languages. Atkins (as quoted by Gouws and Prinsloo (2005: 42)) states that speakers of these languages "have not in their formative years had access to dictionaries of the richness and complexity of those currently available for European languages. They have not had the chance to internalize the structure and objectives of a good dictionary". If user-

friendliness is to be one of the primary considerations in the compilation of the BEDIE, cognisance needs to be taken of the almost complete lack of a dictionary culture and the resulting absence of dictionary consultation skills.

Figure 1: User profile: IE dictionary



4. Planning of the dictionary

The team members involved in the planning of the dictionary deemed it extremely important that the envisaged dictionary should be planned according to sound terminological and lexicographic principles. Taking both the profile of intended target-users and the purpose of the dictionary into account, it was decided that an approach that is user-friendly in the extreme, is the only appropriate one. This has direct implications for both the micro- and macro-structure of the intended dictionary. With regard to microstructure, the value of a rather simplified structure needs to be weighed up against the need to provide the maximum amount of encyclopaedic information. The maximum number of data categories that constitutes the dictionary article is five, including the lemma itself. Compare example (2).

- (2) <lemma> negative type
- <definition> piece of semiconductor doped with impurities, which enables it to donate electrons
- <symbol> N-type
- <cross-reference address> ALSO SEE: positive type
- <synonym> (none)

The ordering typology is strictly alphabetical, rather than thematic as is often the case in LSP dictionaries. However, it has to be kept in mind that one of the functions of a good LSP dictionary is to reveal conceptual relationships existing between different terms. Bergenholtz and Tarp (1995: 199) point out that the alphabetical ordering principle precludes the illustration of conceptual relation-

ships, since concepts occur out of context in an arbitrary order. The user is therefore not given an overview of the conceptual structure of the subject field in question. Although this is a valid point, it needs to be kept in mind that all related concepts do not necessarily share the same headword. A thematic ordering would therefore go some way in revealing conceptual relationships, but does not offer the ultimate solution to the problem. On the other hand, an alphabetic arrangement is practical, fast and familiar to the user and, taking the profile of the target user of the BEDIE into consideration, probably the appropriate ordering principle. However, the exposure of conceptual relationships cannot be sacrificed for the sake of simplicity, therefore alternative strategies need to be employed. For the purpose of this dictionary, mainly two techniques are used. In the first instance, external cross-referencing is employed to explicitly indicate to the user that specific concepts are related. According to Gouws and Prinsloo (2005: 179), an external cross-reference address can either be located elsewhere in the central list, or in a separate text outside the central list. Both these cross-reference types are utilized in the BEDIE, although the second type is aimed not so much at revealing conceptual relationships, but rather at providing additional information from an external source. Compare the articles of *series circuit* and *parallel circuit* in (3) below as an example of cross-referencing to another article in the central lemma list.

- (3) **series circuit** *Circuit in which the components are connected end to end so that the *current* has only one path to follow through the circuit.*
ALSO SEE: *parallel circuit*
- parallel circuit** *Circuit consisting of two *resistors*, connected side by side so that there is more than one path through which *current* can flow.*
ALSO SEE: *series circuit*

An explicit cross-reference marker **ALSO SEE** at the end of both articles refers the user to the related concept. By reading both definitions, the target user can conceptualize the nature of the relationship existing between the two concepts. Secondly, by identifying the correct superordinate concept for all related subordinate concepts, the nature of the relationship between the superordinate and its subordinates is revealed, albeit in a more implicit manner. Compare example (4) in this regard, where a logical relationship exists between the superordinate concept 'metallic element in Group 11 of the Periodic Table (a transition element)' and the two subordinate concepts 'silver' and 'copper'.

- (4) **silver** A silver-white metallic element in Group 11 of the Periodic Table (a transition element), with *atomic number* 47. It conducts *current* and heat very well, even better than *copper* does. SYMBOL: Ag. SEE PERIODIC TABLE p. iv
- copper** A reddish metallic element in Group 11 of the Periodic Table (a

transition element), with *atomic number 29*. It is very much used in electrical work because it conducts *electricity* so well. However, *gold* and *silver* conduct electricity better than copper does. SYMBOL: Cu. SEE PERIODIC TABLE p. iv

To further enhance the user-friendliness of the dictionary, all terms used in any given definition that are treated in the main lemma list of the dictionary, are printed in italics. This convention would of course have to be explained in the user's guide that will form part of the front matter text. The team members also decided to make each terminological definition as extensive as possible, providing the user with more than just a basic paraphrase of meaning. This would compensate to a certain extent for the low level of encyclopaedic competence of the target user. Compare in this regard the definition of the term *current flow* in example (5).

- (5) **current flow** Movement of free *electrons* through a particular material. While the *atoms* of some materials (the *conductors*) give up electrons from their outer orbits freely, other materials hold on to these outer electrons much more tightly. The number of free electrons available is the factor which determines the ease with which *current* will flow through a material. A material with few free electrons will only pass current unwillingly. In other words, it offers opposition to current flow. This opposition to current flow is called *resistance*.

An expanded data distribution structure is envisaged, with the outer text containing amongst others a user's guide, a list of commonly used abbreviations and symbols, frequently used formulae, the Periodic Table as well as a list of academic words such as *define*, *compare*, *identify*, *illustrate*, *motivate*, etc., necessitated by the low level of English proficiency of the intended target users. Where applicable, the user will be referred to the outer texts by means of an explicit cross-reference marker. In (4) above, the marker SEE PERIODIC TABLE (with an indication of the exact page number) in small capitals is used.¹

5. Selection of the lemmas to be treated in the dictionary

According to Bergenholtz and Tarp (1995: 98), metalexicographical literature has thus far paid scant attention to the selection of lemmas for LSP dictionaries. They cite doubt on the part of the metalexicographer about the possibility of theory development as a possible reason for this theoretical constraint. They nevertheless suggest a number of fundamental methodological approaches for the selection of lemmas for inclusion in an LSP dictionary, stating that "quality in practical lexicography includes meticulous, goal-oriented selection of lemmata". The selection process described below has exactly that in mind and represents a novel approach to lemma selection for LSP dictionaries, going

beyond the frequency approach, which seems to have become standard practice for LGP lexicography.

In keeping with developments in the international arena regarding the compilation of LSP dictionaries, it was decided that a corpus-based approach would be followed for this project. The advantages of such an approach have often been stated, one of them being the wealth of encyclopaedic and linguistic information the corpus provides. As Shreve (2001: 773) remarks, the text in which a technical term occurs is an important source of information — not only on its usage, meaning and appropriateness, but also on the relationship it has with other terms. When terms co-occur in a text, conceptual relationships are established. Shreve even suggests that the larger conceptual structure of a special subject field cannot exist unless it has been established by or extracted from a corpus of texts. The proposed procedure for the compilation of the BEDIE has been designed to mine and to maximally utilise any information provided by the corpus. For this particular project, the utilisation of the corpus revolves around two main issues: firstly, the selection of the appropriate terms to be entered into the central lemma list, and secondly, the extraction of definitional information that can be used for the writing of the terminological definitions.

5.1 Compilation of a special-purpose corpus on industrial electronics (IE corpus)

The special-purpose corpus used for the compilation of the BEDIE consists of three textbooks and a separate glossary of terms that form the basis of the curriculum taught at the two FET colleges in the Greater Metropolitan Area of Tshwane referred to earlier. These sources are the following:

- Van Deventer, D.J. *Industrial Electronics N1*. 2000. Cape Town: Maskew Miller Longman.
- Kraft, J. *Industrial Electronics N2*. 2000. Cape Town: Maskew Miller Longman.
- Kraft, J. *Industrial Electronics N3*. 2000. Cape Town: Maskew Miller Longman.
- Kraft, J. *Glossary. Electrical Technology*. 2004. Unpublished glossary compiled by J. Kraft for use in the teaching of Industrial Electronics at the Pretoria West Campus of the Tshwane South College for Further Education and Training.

As the new Curriculum Statement for Industrial Electronics in the FET Band was not yet available when the corpus was being compiled, the compilers were compelled to base the BEDIE on the most recent textbooks used in the teaching

of the curricula at the two colleges, i.e. the IE textbooks for N1, N2 and N3 published in 2000.

As a first step, a small special-purpose corpus on industrial electronics was compiled in which each of the four sources mentioned above was treated as a subcorpus. This was done by scanning the material using *OmniPage* software, then storing the electronic version in ordinary .txt (text) format. This was done after the necessary permission had been obtained from the authors to reproduce the text in electronic format. The sizes of the various subcorpora within the IE corpus in running words (so-called 'tokens') are shown in Table 2.

Table 2: Subcorpora and sources

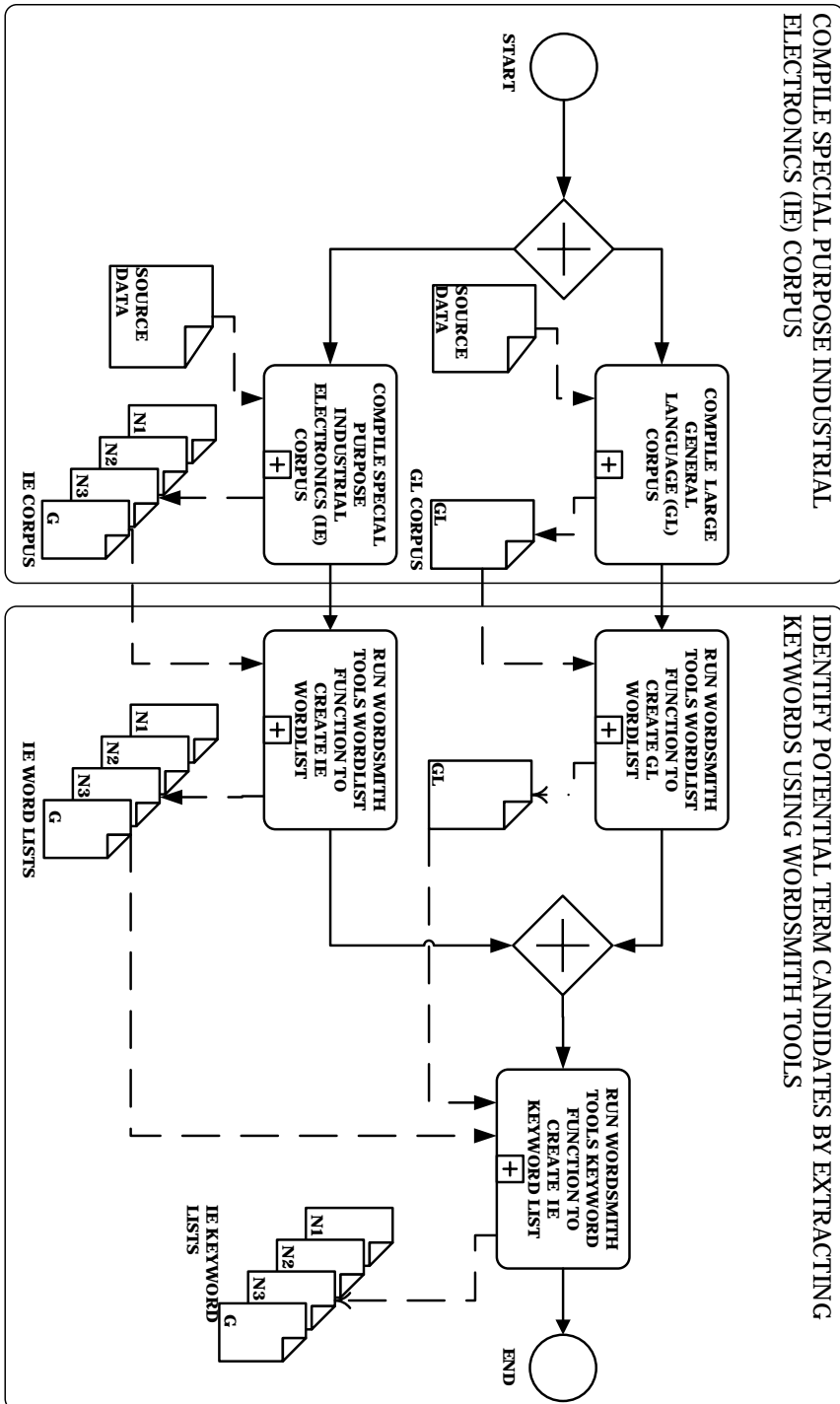
	Source	Size in running words ('tokens')
Subcorpus 1	N1	29 590
Subcorpus 2	N2	23 569
Subcorpus 3	N3	33 699
Subcorpus 4	Glossary	4 699
	TOTAL	91 557

5.2 Semi-automatic extraction of terms

In order to identify and semi-automatically extract the appropriate 500 terms to be treated in the dictionary, the corpus query software *WordSmith Tools* was used, in particular its *WordList* and *KeyWords* functions. In order to identify term candidates for possible inclusion in the BEDIE, the *WordList* function is applied first in order to compile a word list according to frequency of occurrence for each of the four special-field subcorpora *and* for a larger, general-language reference corpus. As a reference corpus, the University of Pretoria English Internet Corpus (PEIC) was used. This electronic corpus consists of 12.5 million English running words (tokens), culled from the Internet by Rachéle Gauton. The *KeyWords* tool is then used to locate and identify keywords in the four special-field subcorpora — the keywords being those words that characterise each of these texts. In order to identify such keywords, the word lists excerpted from each of the small special-purpose corpora (IE subcorpora 1–4) are compared with those of the large general corpus (PEIC). Words with an unusually high frequency in the small specialized corpora are regarded as 'key' and therefore as potential term candidates. This procedure is represented in Diagram 1.

The assumption that underlies the use of the *KeyWords* tool for the purpose of identifying appropriate terms for an LSP dictionary, is that words that appear with an outstanding frequency in a small special-purpose corpus, will probably be the most important or key *terms* in that corpus. It should be noted however, that the *KeyWords* list inevitably contains some 'noise', i.e. items that are obviously not terms. Some manual intervention is therefore needed to re-

Diagram 1: Procedure for the identification of keywords, i.e. potential term candidates



move these items from the list of term candidates. At this stage, the input of a special-field expert is of great value, since it is not always possible for non-specialists to decide whether a specific item is indeed to be regarded as a term in the specific subject field. For this particular project, the assistance of an expert in industrial electronics attached to the physics section of the Foundation Year Programme at the University of Pretoria was sought.²

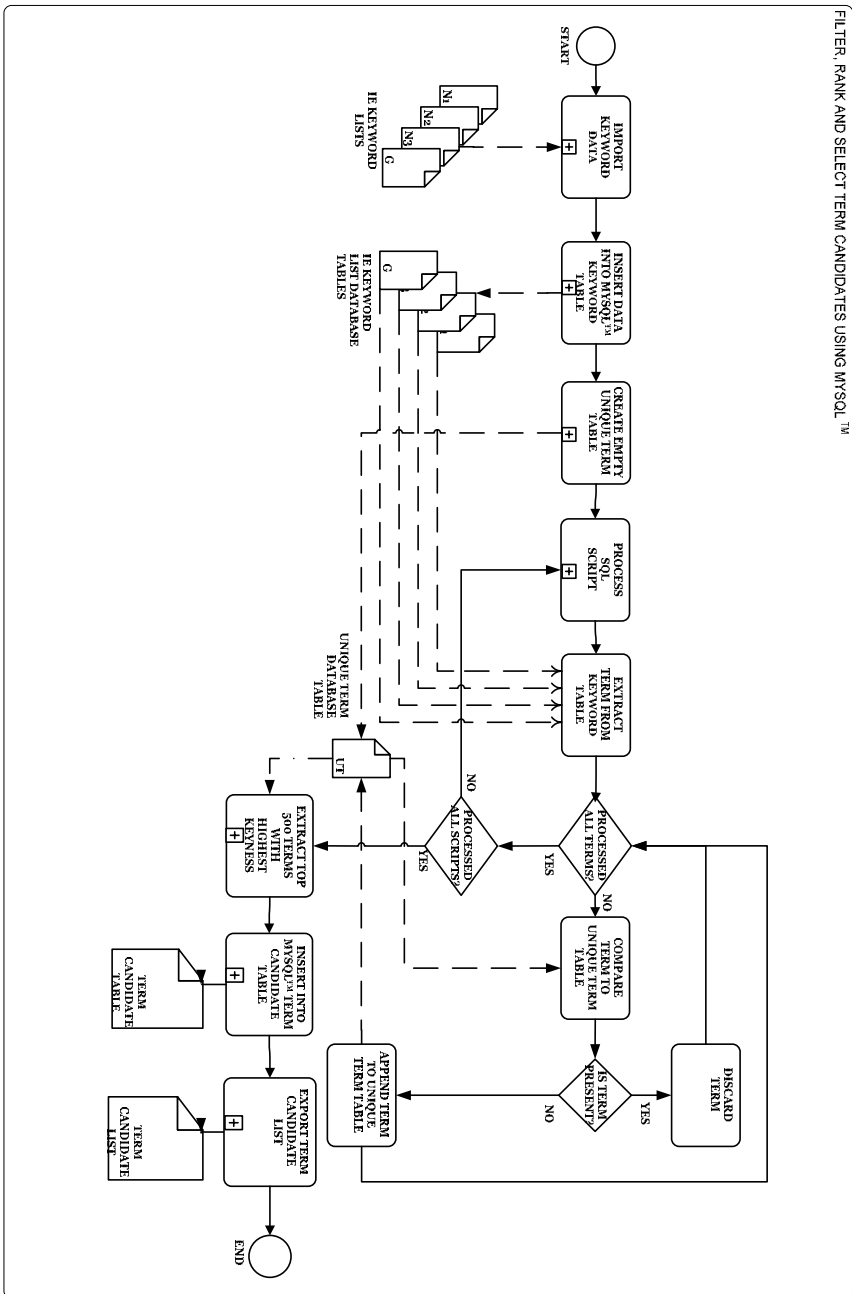
By comparing each of the four IE subcorpora with the large PEIC corpus one-, two-, three-, four- and five-word keywords were subsequently identified for *each* of the subcorpora, representing the study years N1, N2 and N3 as well as the separate glossary of terms (G). (See again Diagram 1 in this regard.)

These keyword lists were then further analysed (through being filtered, ranked and selected) by making use of the open-source database management system *MySQL*TM. Roughly speaking, each of the N1, N2 and N3 KeyWords lists represented a priority list of words/terms from each of three syllabus years that build upon the knowledge base of the learner. The working assumption was that there would be repetition from each preceding year's data in the following year's data. First, the *WordSmith Tools* KeyWords output for each syllabus year was imported into its own database table. An *SQL* query to extract unique terms from the first year's data (i.e. N1) was written and the result set was stored in a new table used to build a consolidated list of unique terms. From the second year's data (i.e. N2), an *SQL* query was constructed to extract unique terms that did not already exist in the unique terms table and the result set was then added to the unique terms table. This process was repeated for the third year's data (N3) as well as for the separate glossary (G).

Once the consolidated list was available in its own database table, it was a simple matter to extract a list of the top 500 terms based on the (high) keyness value of the words as determined by the *WordSmith Tools* KeyWords function. (As stated earlier, words with a high (positive) keyness value are those words with an unusually outstanding (high) frequency when compared with the words in a large (general language) reference corpus). The words/terms in this list were then roughly lemmatized in order not to count e.g. singular and plural forms of the same word as two separate lemmas/headwords. This procedure can be represented visually in Diagram 2.

Subsequently, an attempt was made to validate the 500 identified term candidates by comparing them with keywords extracted from a 58 990 running words (tokens) IE corpus culled from the Internet by Rachéle Gauton (in consultation with Philip Pare). This corpus, although dealing with the field of industrial electronics in general, was compiled in such a way that it also focused on IE texts aimed at learners, although not necessarily South African learners. Again by using *SQL*, the top 500 term candidates list from the special-purpose corpus based on the curriculum taught at the two FET colleges was compared with a list of possible term candidates from the generally available IE Internet corpus to ensure the validity of the terms. On comparing these two lists, it was found that a total of 333 term candidates appear on both lists. A list

Diagram 2: Procedure for filtering, ranking and selecting keywords, i.e. potential term candidates, using MySQL™



FILTER, RANK AND SELECT TERM CANDIDATES USING MYSQL™

containing the remaining terms that appear in the top 500 term candidates list culled from the special-purpose IE corpus, but *not* in the generally available IE Internet corpus, was then presented to the subject specialist for comment. This was done to ascertain the acceptability or not (as the case may be) of these terms for inclusion in the LSP dictionary. This was essentially a cross-check procedure to make sure that terms are not identified as candidates for inclusion in the dictionary should they, for example, not be in use any more and/or incorrect. From the list of term candidates presented to him, the subject specialist selected 147 for inclusion in the BEDIE, resulting in a total of 480 single and multiword term candidates — a number that is close to the original target of 500. This procedure is represented in Diagram 3.

However, what had not been taken into account up to this point, was the fact that the definitions themselves generate more terms for treatment in the dictionary. It is a generally accepted principle that terms used in terminological definitions should also be defined in an LSP dictionary. The definition of the term *capacitance* in (6) serves as an example.

- (6) **capacitance** Property of an object which opposes a change in *voltage*, e.g. the ability of a *capacitor* to store *energy* in an electric field. Capacitance depends upon the distance between the *plates* of a capacitor, as well as on the area of the plates. The larger the plate area, the bigger the *charge* that can be stored. Capacitance is measured in *farads*

Six terms, printed in italics, are used in the definition. Four of these appear on the 480 core term list, but the two terms *plate(s)* and *farad(s)* do not. The recommendation of the subject expert was that these two terms needed to be lemmatised and treated. Thus the definition of the term *capacitance* generated two new terms for inclusion in the lemma list of the dictionary.

A further consideration that comes into play is that of completing paradigms of closely related concepts. The concepts 'proton', 'electron' and 'neutron' are a case in point. The terms *proton* and *electron* both appear in the 480 core list of term candidates, but not the term *neutron*. However, such a close conceptual relationship exists between these three concepts, that it would be a serious oversight not also to lemmatize and fully treat the term *neutron*.

Consequently, an analysis was made of the definitions of the 100 terms with the highest keyness values in the 480 core term list (sorted according to keyness). It was found that a total of 28 extra terms, i.e. terms that do not appear on the core term list, were generated by the definitions of these terms, either because they were used in the definitions or because they formed part of a larger paradigm. At this stage, it is not yet clear whether the 28% increase is a constant that should be taken into consideration by terminologists and LSP lexicographers when selecting terms to be included in an LSP dictionary for which the number of terms to be treated is restricted. It is also not clear if the same kind of increase would present itself when moving down the keyness list,

Diagram 3: Procedure for validating the top 500 term candidates extracted from the special-purpose IE corpus (based on the curriculum taught at the two FET colleges) using MySQL™

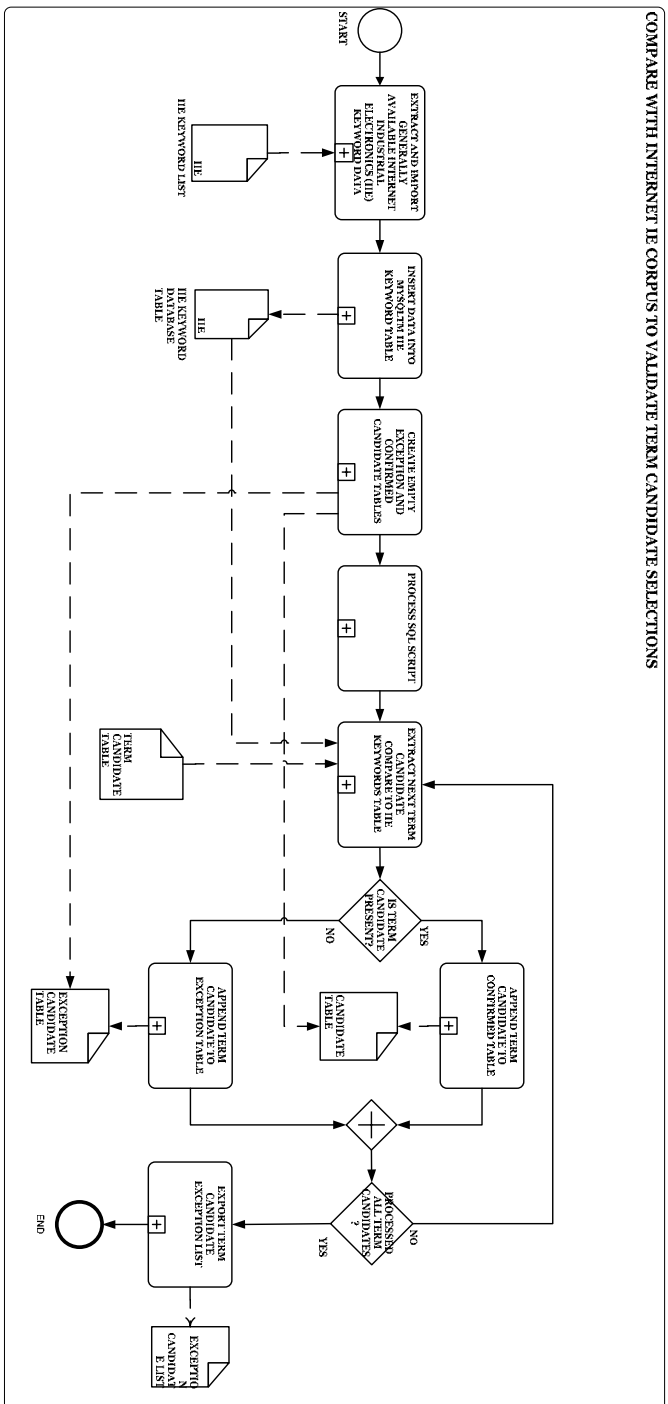
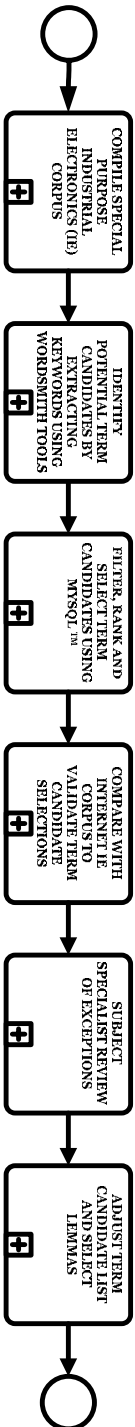


Diagram 4: Procedure followed for the selection of lemmas to be treated in the BEDIE



in other words whether the definitions of the second 100 key terms would generate the same percentage of extra terms. Further investigation is needed before such an increase can be factored in.

In summary, the overall procedure followed in selecting the lemmas to be treated in the dictionary, can be visually represented as in Diagram 4.

6. Semi-automatic extraction of definitional information and compilation of terminological definitions

The 333 term candidates which appear as KeyWords in both the IE corpus based on the curriculum taught at the two FET colleges and the IE Internet corpus aimed at learners, together with the 147 terms (from the top 500 list of term candidates extracted from the special-purpose IE corpus based on the curriculum taught at the two FET colleges) that were added on the recommendation of the subject expert, are regarded as representing the core vocabulary of the IE texts used to build the corpora. As these 480 identified term candidates have been sorted according to their keyness values, definitional information can then be extracted from the text corpora by entering each of the term candidates in order of keyness as a search node in the concordancing function KeyWords in Context (KWIC) of *WordSmith Tools*. The information thus retrieved forms the basis of the terminological definitions, starting at the top 100 key terms. The basic assumption underlying this method is that authors of technical texts very often provide definitional information in these texts. The technique used to semi-automatically retrieve definitional information from electronic texts is described in Pearson (1998: 103). She indicates that either lexical or syntactic devices signal the presence of definitional information in texts. A typical syntactic device is the use of the copula, whereas typical lexical markers are expressions such as *is/are called*, *is/are known as*, etc. Since the IE corpus that is used for the compilation of the BEDIE is untagged, mainly lexical markers were relied upon for the identification of definitional information. Even so, since the IE corpus is relatively small (91 557 tokens), manual scanning of all KWIC lines generated for a specific search word is not a huge undertaking, thus the search for definitional information is not restricted to the identification of information marked by lexical items. Compare the example in (7) below, representing a sample of KWIC lines in which the term *electromagnet* is the search node.

(7) Sample of KWIC lines with *electromagnet* as search node

10. Such a device is known as an **electromagnet** and is used extensively in the Magnetism remaining in the core of an **electromagnet** after the coil current is removed. correct polarity to prevent breakdown. **electromagnet** A coil of wire usually wound on a A relay makes use of an **electromagnet**. When current flows through the coil that is tripped or activated by use of an **electromagnet**. magnetic coil: Spiral of a conductor

By making use of the Grow facility of *WordSmith Tools*, more context surrounding the search node can be revealed, as shown in (8).

- (8) More context surrounding the search node *electromagnet* as revealed by the Grow function

As iron provides a better path (higher permeability) for the lines of force than air, the strength of the magnetism is much greater as shown in figure 6.10. Such a device is known as an **electromagnet** and is used extensively in the manufacture of electronic equipment such as relays, doorbells, buzzers and circuit breakers.

remenance Amount a material remains magnetized after the magnetizing force has been removed. residual magnetism Magnetism remaining in the core of an **electromagnet** after the coil current is removed. resistance Symbolized "R" and measured in ohms.

The oxide acts as the dielectric for the capacitor. Electrolytic capacitors are polarized and so must be connected in correct polarity to prevent breakdown. **electromagnet** A coil of wire usually wound on a soft iron or steel core. When current is passed through the coil a magnetic field is generated. The core provides an easy path for the magnetic lines of force. This concentrates the field in the core.

In order to use the coil as a controlled electromagnet, it should be wound on a core that retains little of its magnetism. A relay makes use of an **electromagnet**. When current flows through the coil, the iron core is magnetised.

magnet Body that can be used to attract or repel magnetic materials. magnetic circuit breaker Circuit breaker that is tripped or activated by use of an **electromagnet**. magnetic coil Spiral of a conductor which is called an electromagnet. magnetic core Material that exists in the center of the magnetic coil to either physically support the windings (non-magnetic material) or to concentrate the magnetic flux (magnetic material).

Based on the definitional information retrieved from the KWIC lines, the definition of the term *electromagnet* in (9) was formulated.

- (9) **electromagnet** Magnet which consists of a *coil* of wire that is usually wound on a soft *iron* or *steel* core. When *current* is passed through the coil a *magnetic field* is generated. The core provides an easy path for the *magnetic lines of force*. Electromagnets are widely used in the manufacture of electronic equipment such as doorbells, buzzers, relays and *circuit* breakers.

After the definitions based on information retrieved from the KWIC lines have been formulated, these are submitted to the subject field expert for final checking of contents. At this stage of the compilation process, the input of the subject-field expert is essential, since the lexicographer — not being an expert in industrial electronics — has no way of judging whether the salient features of a particular concept had indeed been thrown up by the corpus. Where necessary, definitions are then revised. Compare example (10) in this regard, where (a) is the definition compiled by the lexicographer, based on definitional information extracted from the corpus, and (b) is the definition as revised by the expert.

- (10) (a) **tunnel diode** *Semi-conductor* device that will exhibit a negative *resistance* between the values of 0.2 *volts* and 0.4 *volts* when *forward-biased*. It has a low peak *tunnel current* and is used as a low-voltage *rectifier* biased in the reverse direction.
- (b) **tunnel diode** *Semi-conductor* device that has a negative *resistance* between the values of about 0.2 *volts* and 0.4 *volts* when *forward-biased*. It is usually used as a high-speed switch and in high-frequency *oscillator circuits*. These are *circuits* which are used to produce high-frequency *alternating currents*. It was invented by Leo Esaki in 1958.

In cases where no definitional information can be retrieved from the corpus, definitions are supplied by the subject-field expert.

One of the main advantages of the procedure described above is the fact that very little claim is laid to the knowledge of the terminologist/LSP lexicographer of the particular subject field. The retrieval of definitional information from the corpus is to a large extent a mechanical process. Furthermore, the demands made on the time and input of the subject-field expert are also lessened — something which might make experts more inclined to take part in such projects.

7. Translation of terms and definitions

After final checking by the subject-field expert, the list of lemmas and their definitions are handed over to the translator for rendering them into Northern Sotho. As indicated in the introduction, a discussion of the translation process does not fall within the ambit of this article and will therefore not be addressed here.

8. Conclusion

In this article, a possible approach to the compilation of an LSP dictionary was discussed by presenting the planning and design of the BEDIE as a case in point.

The main contribution this article makes to the field of (corpus) lexicography, and specifically to the discipline of (corpus) LSP lexicography, is that it details a highly effective and functional approach towards the selection of lemmas and subsequent extraction of definitional information for an LSP dictionary in the most labour-efficient manner (i.e. computer aided as opposed to manual) by using various functionalities of (general purpose) software such as *WordSmith Tools* and *MySQL™*.

Notes

1. The page number cited in this example is a fictional one, added for illustrative purposes.
2. We would like to express our thanks to Phillip Pare, who unstintingly shared his time and expertise with us. Without his input, this project would not have been possible.

Bibliography

- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Gauton, R., E. Taljard, T.A. Mabasa and L.F. Netshitomboni.** Forthcoming. Translating Technical (LSP) Texts into the Official South African Languages: A Corpus-based Investigation of Translators' Strategies. (To be submitted to *Language Matters*).
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Kraft, J.** 2000. *Industrial Electronics N2*. Cape Town: Maskew Miller Longman.
- Kraft, J.** 2000. *Industrial Electronics N3*. Cape Town: Maskew Miller Longman.
- Kraft, J.** 2004. *Glossary. Electrical Technology*. Unpublished glossary used in the teaching of Industrial Electronics at the Pretoria West Campus of the Tshwane South College for Further Education and Training.
- MySQL™ Version 5.0.1. Available from <http://www.mysql.com>.
- OmniPage. Available from <http://www.nuance.com>.
- Pearson, J.** 1998. *Terms in Context*. Amsterdam: John Benjamins.
- Scott, M.** 1999. *WordSmith Tools Version 3*. Oxford: Oxford University Press. Also available from <http://www.lexically.net/wordsmith/index.html>.
- Shreve, G.M.** 2001. Terminological Aspects of Text Production. Wright, Sue Ellen and Gerhard Budin (Eds.). 2001. *Handbook of Terminology Management. Volume 2. Application-Oriented Terminology Management: 772-787*. Amsterdam: John Benjamins.
- Van Deventer, D.J.** 2000. *Industrial Electronics N1*. Cape Town: Maskew Miller Longman.
- Webb, V.N.** 2005. *Report on the Project Language, Educational Effectiveness and Economic Outcomes*. Submitted to the Swiss Development Agency. Pretoria: University of Pretoria, CentRePoL.

Lexicography in the Information Age*

Sven Tarp, *Centre for Lexicography, Aarhus School of Business, Aarhus, Denmark (st@asb.dk)*

Abstract: Lexicography has much to contribute to the information and knowledge society. The very essence of lexicography is its capacity to provide quick and easy access to data from which information needed by different types of users in different types of social situations can be retrieved. In this respect, the article discusses the concepts of information and knowledge related to lexicographic theory. On this basis, it argues that lexicography, if it concentrates on its core capacity, has the potential to be projected beyond the limits of known dictionaries. It proposes that besides the communicative and cognitive user situations that have already been dealt with in lexicographic literature, a third situation, the so-called operational situation, should also be added to the situations relevant to lexicographic theory. In this regard, it points out that other reference works than those traditionally discussed in lexicographic theory should be added to the list of utility tools that may benefit from a lexicographic approach.

Keywords: LEXICOGRAPHY, LEXICOGRAPHIC TOOLS, INFORMATION NEEDS, INFORMATION AND KNOWLEDGE SOCIETY

Opsomming: Leksikografie in die inligtingsera. Leksikografie het baie om by te dra tot die inligtings- en kennisgemeenskap. Die werklike essensie van die leksikografie is sy vermoë om vinnige en maklike toegang te verleen tot gegewens waarvan inligting benodig deur verskillende soorte gebruikers in verskillende soorte maatskaplike situasies verkry kan word. In hierdie verband bespreek die artikel die begrippe inligting en kennis met betrekking tot die leksikografiese teorie. Op grond hiervan voer dit aan dat die leksikografie, indien dit op sy kernvermoë konsentreer, die moontlikheid besit om verby die grense van bekende woordeboeke gevoer te word. Dit stel voor dat, benewens die kommunikatiewe en kognitiewe gebruikerssituasies wat reeds in die leksikografiese literatuur behandel is, 'n derde situasie, die sogenaamde operasionele situasie, by die situasies tersaaklik vir die leksikografiese teorie gevoeg behoort te word. In hierdie verband toon dit aan dat ander naslaanwerke as dié wat tradisioneel in die leksikografiese teorie bespreek word, ook by die lys nutsgereedskap wat uit 'n leksikografiese benadering mag voordeel trek, gevoeg behoort te word.

Sleutelwoorde: LEKSIKOGRAFIE, LEKSIKOGRAFIESE GEREEDSKAP, INLIGTINGS-BEHOEFTES, INLIGTINGS- EN KENNISGEMEENSAP

* 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.

It is frequently stated that we are living in a so-called information and knowledge society. Such statements are often highly superficial, for example when this kind of society is considered as synonymous with a post-industrial society where industry allegedly should have given way to something else. With this claim, industry is mainly defined according to some historically ephemeral features, i.e. its temporary fordist features, and not according to its more general role as the most important organisational form of human production even today. But apart from such examples where secondary features are transformed into primary ones, it cannot be ignored that knowledge and information have become increasingly important components of present-day society in general and post-fordist industry in particular. In this respect, it is acceptable to speak of an information and knowledge society.

But what has all this to do with lexicography?

When an abstraction is made from the dictionary form, i.e. from both the traditional printed dictionary and the more advanced electronic one, what is left is a utility tool whose essential characteristic is its capacity to provide quickly accessible data from which information can be retrieved. And these data are furthermore selected and prepared in order to satisfy the information needs, not of a specific type of users in general, but of a specific type of users finding themselves in a specific type of extra-lexicographic situation where such needs may emerge. This is the very essence of lexicography which can be projected far beyond the traditional limits of known dictionaries. This contribution will analyse this idea and outline some of its interesting perspectives.

Lexicographic needs

Although Wiegand (1988) was the first to state it explicitly, the idea that dictionaries are utility tools which should be designed for their users is not new. For instance, at the end of a classic conference on lexicography in 1960, Householder (1967: 279) made an important recommendation that has been quoted repeatedly, not least in English-language lexicography:

Dictionaries should be designed with a special set of users in mind and for their specific needs.

At an abstract level, it is hard to disagree with this recommendation. But at least two fundamental factors are missing in terms of the potential users' needs which in a certain way are left in a vacuum. In the first place, users do not have specific needs unless they are related to specific types of situations. Consequently, it is not enough to define which types of users have which needs, but the types of social situations in which these needs may arise should also be described. Not all social situations are relevant to lexicography, but only situations in which needs may arise that can be satisfied by consulting dictionaries or other lexicographic tools.

Secondly, even in these relevant situations, not all needs are lexicographi-

cally relevant. If a translator, for example, becomes hungry while practising his/her profession, this need for food is completely irrelevant to lexicography. Hence, the general type of needs that are lexicographically relevant should also be defined and later subdivided according to the various types of relevant user situations and user profiles.

As it has already been mentioned, the only types of needs that can be met by dictionaries and other lexicographic tools are the needs for information. Users consult dictionaries in order to obtain information which they can then use for a great variety of purposes, e.g. to solve communication problems or to add to their existing knowledge. However, dictionaries themselves do not contain information, but only lexicographically selected and prepared data from which the users may or may not be able to retrieve the needed information. Wiegand (1998, 2000, 2002) introduced this important distinction between what is contained in the dictionaries and what is retrieved from them through a cognitive process during a successful consultation. His distinction is based on a philosophical approach according to which knowledge — and information — cannot exist separated from the human brain.

It could be argued that such a philosophical discussion is irrelevant to lexicography and that lexicographers would surely have different opinions on a crucial question of this kind. However, even if some lexicographers disagree with Wiegand's basic premises due to their extra-lexicographic world view, it would be very useful to maintain the distinction between the lexicographic data contained in the dictionary and the information extracted from these data. If a dictionary user, for instance, wants to know something about geology and consults a Chinese encyclopaedic dictionary of geology, the information that can be retrieved depends on the user's proficiency level in Chinese and not on the correctness of the lexicographic data. Although these data are not prepared for non Chinese-speaking users, who will probably not be able to extract any useful information at all, this does not mean that it is impossible for the anticipated type of users to retrieve the needed information. Besides, such specialised dictionaries of geology should not only be conceived for speakers of a specific language, but they should also be adapted to the users' scientific level, i.e. whether they are experts, semi-experts or lay people. A lay person, for instance, will probably not be able to retrieve any useful information from a dictionary designed for an expert although all the necessary data are included in the dictionary, but in such a complex way that these are almost incomprehensible without a specialised knowledge of geology. Thus, lexicographic data is something different from the information retrieved from these data.

Data and quick accessibility

There are various written and oral sources from which human beings can acquire information for some purpose or other. Among the written sources, the most important are newspapers, journals, magazines, books, text books, inter-

net-based texts and dictionaries. In this respect, no information that cannot be retrieved from these sources is obtained by the consultation of dictionaries. For many purposes, some of these sources, especially text books, may be even more appropriate than dictionaries. Nevertheless, dictionaries have some specific benefits which make them far more suitable for other purposes.

- Firstly, dictionaries furnish lexicographically selected and prepared data which are, at least in theory, especially adapted not only to specific types of users (something which they have in common with other written sources), but also to specific types of user situations (where they generally have few rivals).
- Secondly, and even more importantly, the relevant data included in dictionaries are made easily and quickly accessible through different kinds of prepared and foreseen access routes (cf. Bergenholtz and Gouws 2007, and Tarp 2008).

Accessibility is a key concept in any lexicographic theory maintaining to be user-oriented. However, it could be claimed that any data included in any text is accessible to anyone who takes the time to read the text through from beginning to end. Other tools, especially browsers connected to the internet, also provide easy access to relevant data, but as the search on the internet in most cases produces much redundant material in need of further processing, it can hardly be considered quick.

Hence, one of the really distinctive features of dictionaries and other lexicographic tools is that they provide quick and easy access to the specific types of data that can cover a specific type of user's specific types of information needs in a specific type of extra-lexicographic situation. In this respect, lexicography is highly relevant to a society where information is increasingly important to all its members in their professional and daily life and which even defines itself as an information and knowledge society.

The relation between information, knowledge and communication

There is a large variety of uses and definitions of the word *information*. In Wikipedia (June 15, 2007) the following short explanation of *information* can be found:

Information is the result of processing, manipulating and organizing data in a way that adds to the knowledge of the receiver. In other words, it is the context in which data is taken. Information as a concept bears a diversity of meanings, from everyday usage to technical settings. Generally speaking, the concept of information is closely related to notions of constraint, communication, control, data, form, instruction, knowledge, meaning, mental stimulus, pattern, perception, and representation. Many people speak about the Information Age as the advent of the Knowledge Age or knowledge society, the information society, and information technologies, and even though informatics, information science and

computer science are often in the spotlight, the word "information" is often used without careful consideration of the various meanings it has acquired.

It is logical that the definition of central concepts frequently varies from science to science, and even from theory to theory within the same science, as they all have their own specific approaches to these concepts. The same applies to lexicography which needs its own definitions, not only of *information*, but also of *knowledge*, *communication* and other important concepts. In this respect, the lexicographic function theory agrees with the above definition that "information is the result of processing [...] data", but it does not agree that the information is necessarily something processed, manipulated and organised "in a way that adds to the knowledge of the receiver". The function theory sees information as a kind of "proto-knowledge" with the potential to be transformed into knowledge, but also to be used for other purposes. In this regard, knowledge is considered as information stored in the brain — through cognitive processes which is not the study field of lexicography — for subsequent use.

There are two main reasons for this clear distinction between information and knowledge in the lexicographic approach. Firstly, dictionary usage shows that users often, within a period of some hours or days, consult the same entry two or three times in order to satisfy their need for the same specific information. In such cases, the information originally retrieved from the data contained in the dictionary may have served to meet the users' immediate needs, but it has obviously not been completely and satisfactorily stored in their memory for subsequent use.

Secondly, not all dictionary consultations are aimed at adding to the users' knowledge. For instance, when it is a question of problems or doubts arising during the translation of texts from one language into another, or during other communication processes, the purpose of the dictionary consultation is exclusively to find answers to these problems and doubts. In such types of user situations, the information retrieved from the dictionary is needed as a kind of "trouble-shooter" that may have only an ephemeral existence. Although it cannot be excluded that the information in such cases is eventually also stored in the memory, it is important to clarify that this is not the direct purpose of the dictionary consultation, but rather a sort of fortunate by-product of this consultation.

In this respect, there is a complex dialectic relation between information, knowledge and communication (cf. Tarp 2005). The very process of retrieving information from a dictionary is essentially a delayed communication between the original compiler(s) of the dictionary and the user. The information needed for immediate communication purposes may eventually be stored in the memory as knowledge, and the information retrieved with the direct purpose of adding to the user's existing knowledge may likewise be used for subsequent communication processes. In order not to become lost in this complex relationship, it is important always to focus on the original type of information needs that gave rise to the consultation process, and this basically concerns the type of user situation.

Cognitive user situations

The most thoroughly treated and analysed user situations in theoretical literature are the communicative ones, e.g. text production, reception, translation, marking, revision and proofreading. This is probably due to the fact that theoretical lexicography has so far been mostly dominated by scholars with a linguistic background. Whatever the reason, it is of vital importance for lexicographic theory to advance towards a similar understanding of the so-called cognitive user situations which have up till now been defined in contrast to the communicative user situations as those where users for one reason or another want or need to acquire knowledge about a specific phenomenon. This want or need to gain new knowledge can arise in countless social situations, for instance:

- (1) while reading: the sudden need to acquire additional encyclopaedic knowledge in order to understand the text;
- (2) while writing: the need to know more about a given topic in order to finish a text;
- (3) during discussions with other people: the need to clarify a specific issue;
- (4) during processes in the subconscious: the sudden desire to examine something;
- (5) during dictionary consultation: the desire to know more about a specific topic;
- (6) during preparation for specialised translation and interpretation tasks: the need to learn more about the subject field in question;
- (7) in relation to a teaching programme or a course of study: the need to know more about a specific subject field;
- (8) ...

The options are unlimited. However, as most people will be familiar with at least some of the above-mentioned user situations, it is possible even on the basis of this limited selection to deduce some preliminary conclusions. Inspired by the distinction which Hausmann (1977) made between punctual and global issues in regard to dictionary consultation, a fundamental distinction can be made between *sporadic* and *systematic* user situations. Of the above seven situations, the first five are isolated situations without any clear relation to previous or subsequent situations giving rise to dictionary consultation, while the two last ones are situations where the potential dictionary user is making — or is supposed to make — a systematic effort to gradually acquire knowledge about a specific subject field and where a specific consultation in one or the other way is related to previous or subsequent consultations. This distinction between systematic and sporadic situations has important consequences for the design of a dictionary conceived to satisfy the potential users' needs, because the needs arising from a systematic study of a given subject field will require a more sophisticated data distribution structure with synopsis articles and special sections containing a global introduction to the subject field in question as

well as a highly developed system of cross-references or -links allowing users to navigate in the dictionary in order to make full use of its data.

In addition to the distinction between sporadic and systematic user situations, there is a second important conclusion that can be deduced from the above-mentioned user situations. In the systematic situations, it is evident that the purpose of the dictionary consultation is *to add to the users' existing knowledge* (cf. the quoted explanation from Wikipedia). However, at least in two of the other situations, (2) and (3), the direct purpose of the consultation is *to meet an immediate knowledge need* and whether or not the information is added to the user's existing knowledge, i.e. stored in the memory, is a secondary question for the user. When a journalist, for instance, is writing an article about Napoleon and has to inform the readers about the date of birth of this statesman, he/she may then consult a dictionary with the immediate purpose of finding this specific information which he/she may never need again and may probably forget in a few minutes.

Finally, there is a third interesting conclusion that can be deduced from the seven user situations mentioned. In situations (1), (2) and (6), the needs for knowledge are directly related to a current or future communication process and, as such, they could be included among the needs arising during this process. However, from a lexicographic point of view, i.e. in terms of the possible consequences for dictionary concepts, it is necessary to distinguish between the users' inadequate linguistic and other communicative *skills* on the one hand and the possible lacunae in their general or specialised encyclopaedic *knowledge* on the other.

For instance, if a person is reading the poem "Lullaby of the Onion", written by the Spanish poet Miguel Hernandez and dedicated to his only son, and does not know that it was composed in jail immediately after the Spanish Civil War as a reply to a letter from his wife telling him she was surviving on bread and onions, then this person may find symbolisms in the poem that do not actually exist. In this respect, additional encyclopaedic knowledge is necessary to understand the general context in which the poem was written and, thus, the poem itself. As regards text production, a distinction should be made, in an analogous way, between *what* is to be communicated and *how* it should be communicated where the latter concerns the text producer's linguistic and communicative skills while the former concerns his/her general or specialised encyclopaedic knowledge. The same holds true for translation where translators frequently need knowledge about either the general context or the subject field of the text to be translated. This knowledge is of a different nature from the translators' general translation skills, including their proficiency level in terms of specialised vocabulary.

Those engaged in the three mentioned types of communication (1) may have the necessary knowledge in advance, (2) may acquire it through a conscious study before starting the communication process, and (3) may acquire it when their lack of knowledge becomes problematical during this process. It is

evident that both situations (2) and (3) are part of the communication process in the broad sense of the word. However, if situation (2), i.e. a conscious study of a given subject field, occurred without relation to a communicative situation, it would clearly be considered a cognitive situation. On the other hand, the only real difference between situations (2) and (3) is temporal, i.e. between the different moments when the needs occur which give rise to a lexicographic consultation. In this regard, the two types of situations are, in fact, cognitive situations related to communicative situations and could be called *communication-related cognitive sub-situations*.

Although some lexicographers may consider this discussion too extreme and academic, the theoretical conclusions drawn from it are of great relevance for lexicographic practice in as much as the user needs arising in the different situations require different lexicographic solutions and in many cases also dictionary compilers with completely different skills.

New user situations

So far, the lexicographic function theory has discussed two fundamental types of user situations, the cognitive ones and the communicative ones. However, in theoretical literature there are some sporadic references to reference works such as handbooks, manuals and how-tos. These references have not been properly discussed in any known contribution, but it is easy to see that many of these reference works, especially the how-tos, are not designed to provide assistance in either cognitive or communicative situations, but to give instructions on how to proceed in specific situations, e.g. in relation to the operation of machines and other instruments. These situations which may be called *operational situations* have so far not been integrated into the function theory or any other lexicographic theory, for which reason the above-mentioned reference works may not be considered lexicographic tools either.

However, although it is evident that not all handbooks, manuals or how-tos are designed as lexicographic products, there is no reason why they should not be conceived and considered as such. They all have something fundamental in common with traditional lexicographic products, i.e. they are tools conceived to be consulted by specific types of users in order to satisfy specific types of information needs in specific types of social situations. In this respect, they are expected to provide quick and easy access to the relevant data from which the needed information can be retrieved.

It is evident that a lexicographic theory focusing on quick and easy access to data from which specifically user- and situation-adapted information can be extracted has much to contribute to the design of a new generation of improved handbooks, manuals, and how-tos. Moreover, to this list of reference works should be added other tools such as user guides which are frequently not only used to be read from beginning to end, but also — and even mainly — for punctual consultation in order to acquire information that can be trans-

formed into instructions on how to proceed in operational situations very similar to those where handbooks, manuals and how-tos may provide assistance. All these reference works traditionally used in relation to operational situations could benefit from a lexicographic approach which analyses the types of information needs that may arise for specific types of users in the various types of operational situations in order to prepare new types of lexicographic tools with quick and easy access to the relevant data.

In this way, the operational situation should be added to the other fundamental user situations included in the lexicographic function theory. Two-legged lexicography has acquired its third leg wherever it is placed.

Lexicography at a crossroads

After more than 4 000 years of existence, lexicography finds itself at a crossroads. It will either throw itself into the future with all its renewed theoretical and practical strength or it will become circumvented and confined to the sphere of traditional dictionaries considered as products of applied linguistics. There are three reasons for this challenging situation: firstly, the theoretical efforts carried out during the last decades which have revealed the real essence of lexicography as the capacity of providing information; secondly, the advent of the information and knowledge society; and thirdly, the development of the new computer and information technologies.

In order to meet the new challenges, lexicography must, on the one hand, focus on its core knowledge and, on the other, project itself far beyond its traditional limits. It should further study and develop its knowledge about user situations, user profiles, information needs, data selection and presentation, and quick and easy data accessibility. It should promptly and totally adapt itself to the new technologies and it should, among others, explore the possibilities of performing user needs adapted searches on the internet combining traditional static data with dynamic data made available through the internet.

There is no doubt that due to its focus on information needs related to specific types of users and user situations, lexicography is in a position to furnish essential contributions to the new information sciences. In this respect, it may be considered whether lexicography should not change its heavily connotative name and find a new name which corresponds far more to its real essence. Perhaps *infoLOGY* or *informology* may qualify as candidates in this search for lexicography's present-day and future identity.

Literature

- Bergenholtz, Henning and Rufus H. Gouws.** 2007. The Access Process in Dictionaries for Fixed Expressions. *Lexicographica* 23. [Forthcoming.]
- Hausmann, Franz Josef.** 1977. *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.

- Householder, Fred W.** 1967. Summary Report. Householder, F.W. and S. Saporta (Eds.). 1967. *Problems in Lexicography*: 279-282. Bloomington: Indiana University.
- Tarp, Sven.** 2005. The Pedagogical Dimension of the Well-Conceived Specialised Dictionary. *Ibérica. Journal of the European Association of Languages for Specific Purposes* 10: 7-21.
- 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. Tübingen: Max Niemeyer. [Forthcoming.]
- Wiegand, Herbert Ernst.** 1988. Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischem Wissen. Munske, H.H. et al. (Eds.). 1988. *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern*: 729-790. Berlin/New York: Walter de Gruyter.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin/New York: De Gruyter.
- Wiegand, Herbert Ernst.** 2000. Wissen, Wissenrepräsentationen und Printwörterbücher. Ulrich, H., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, 8-12 August 2000*: 15-38. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Wiegand, Herbert Ernst.** 2002. Wissen in der Sprachlexikografie. Ein Plädoyer für einige immer noch notwendige Differenzierungen. Ezawa, K., W. Kürschner, K.H. Rensch and M. Ringmacher (Eds.). 2002. *Linguistik jenseits des Strukturalismus. Akten des II. Ost-West-Kolloquiums Berlin 1998*: 265-281. Tübingen: Gunter Narr Verlag.
- Wikipedia. The Free Encyclopedia.* <http://en.wikipedia.org>.

Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform

Herbert Ernst Wiegand, *Germanistisches Seminar, Universität Heidelberg,
Heidelberg, Bundesrepublik Deutschland*
(herbert.ernst.wiegand@gs.uni-heidelberg.de)

Zusammenfassung: Nach der Einführung der terminologischen Adjektive *adkurrent*, *inkurrent*, *innenkurrent* und *exkurrent* sowie einer Reihe von lexikographietheoretischen Begrifflichkeiten werden Zugriffspfade z.T. anhand von Beispielen erklärt und zahlreiche Zugriffspfadtypen unterschieden. Eine grundlegende Unterscheidung ist die zwischen äußeren und inneren Zugriffspfaden; bei den äußeren Zugriffspfaden werden einfache und zusammengesetzte äußere Zugriffspfade, bei den einfachen Zugriffspfaden mediostrukturelle und nichtmediostrukturelle äußere Zugriffspfade differenziert; weiterhin werden bei den mediostrukturellen binnenkurrente, verzeichnisexkurrente und wörterbuchexkurrente äußere Zugriffspfade unterschieden, sowie bei den nichtmediostrukturellen Zugriffspfaden die adkurrenten, binnenkurrenten und wörterbuchexkurrenten äußeren Zugriffspfade. Der Beitrag stellt insgesamt begriffliche Differenzierungen und eine dazugehörige Terminologie zur Verfügung, mit deren Hilfe die Datenakzessivität von Printwörterbüchern systematisch evaluiert und geplant sowie ihr mediostrukturelles Profil systematisch dargestellt werden kann.

Stichwörter: ADKURRENTER ÄUSSERER ZUGRIFFSPFAD, AKZESSIVE ANGABE, ÄUSSERE ZUGRIFFSSTRUKTUR, ÄUSSERES ZUGRIFFSTEXTELEMENT, BINNENKURRENTER ÄUSSERER ZUGRIFFSPFAD, DATENAKZESSIVITÄT, EINTRAGSEINGANG, EXKURRENTER ÄUSSERER ZUGRIFFSPFAD, EXTERNE VERWEISBEFOLGUNGSHANDLUNG, EXTERNE ZUGRIFFSHANDLUNG, INKURRENTER ÄUSSERER ZUGRIFFSPFAD, INNERE ZUGRIFFSSTRUKTUR, MEDIOSTRUKTURELLE ORIENTIERUNG, ORDNUNGSSTRUKTUR, VERWEISVERMITTELNDES TEXTSEGMENT, ZUSAMMENGESETZTER ÄUSSERER ZUGRIFFSPFAD

Abstract: **About Access Routes in Printed Dictionaries. A Contribution towards the Intersection of Use Procedures and Dictionary Form.** Following the introduction of the terminological adjectives *adcurrent*, *incurrent*, *inner-current* and *excurrent* as well as a series of lexicographic-theoretical concepts, access routes are explained, partly by using examples, and a variety of types of access routes are distinguished. A basic distinction is that between outer and inner access routes. Outer access routes are divided into single and complex ones with single access routes being divided into mediostuctural and non-mediostuctural outer access routes. Mediostuctural outer access routes are divided into inner-current, table of contents excur-

rent and dictionary excurrent outer access routes whilst non-mediostructural access routes are divided into adcurrent, inner-current and dictionary excurrent outer access routes. This article presents conceptual differentiations and a relevant terminology with which the data accessibility of printed dictionaries as well as their mediostructural profile can be evaluated and planned systematically.

Keywords: ACCESSIBLE ITEM, ADCURRENT EXTERNAL ACCESS ROUTE, COMBINED OUTER ACCESS ROUTE, CROSS-REFERENCE ENABLING TEXT SEGMENT, DATA ACCESSIBILITY, ENTRY ENTRANCE, EXCURRENT OUTER ACCESS ROUTE, EXTERNAL CROSS-REFERENCE FOLLOW-UP PROCEDURE, INCURRENT OUTER ACCESS ROUTE, INNER ACCESS STRUCTURE, INNER-CURRENT OUTER ACCESS ROUTE, MEDIOSTRUCTURAL ORIENTATION, ORDERING STRUCTURE, OUTER ACCESS PROCEDURE, OUTER ACCESS STRUCTURE, OUTER ACCESS TEXT ELEMENT

1. Zur Schnittstellen- und Pfadmetaphorik

Unter einer *Schnittstelle* wird in unterschiedlichen Bereichen etwas Verschiedenes verstanden. Die Schnittstelle zwischen einem Benutzer-in-actu und dem Printwörterbuch, das er benutzt, ist eine Menge von lexikographischen Schnittstellenelementen, die bestimmte Benutzungshandlungen ermöglichen bzw. erforderlich machen. Bei einem Printwörterbuch als einem statischen Informationssystem sind alle Schnittstellenelemente, die hier interessieren, Komponenten der Wörterbuchform, und zwar besonders solche, die externe Zugriffshandlungen ermöglichen sowie externe Verweispfandhandlungen erforderlich machen und damit die externe lexikographische Datenakzessivität garantieren (vgl. Wiegand 2005). Dies sind vor allem die äußeren Zugriffsstrukturen in ihren unterschiedlichen Ausprägungen sowie die verweisvermittelnden Textsegmente in ihren vielseitigen Formen (vgl. dazu Wiegand 2002). Beim Benutzer ist die Beherrschung der allgemeinen Wörterbuchbenutzungspraxis (*sensu* Wiegand 1998: 364ff) die wesentliche Voraussetzung dafür, dass die Schnittstelle funktioniert.

In einer bestimmten Perspektive können die Ausführungen von Benutzungshandlungen als Bewegungen eines Benutzers-in-actu zu einem Wörterbuch und innerhalb eines Wörterbuchs beschrieben werden, die jeweils bestimmte individuelle Pfade etablieren. Liest Werner beispielsweise einen Text und versteht ein Wort W_1 nicht, so dass er daraufhin zu einem Wörterbuch Wb_1 greift und eine externe Zugriffshandlung erfolgreich ausführt, dann kann man z.B. sagen: Werner hat einen Zugriffspfad etabliert, auf dem er von W_1 zu Wb_1 gelangt ist. Erschließt Werner in einem Wörterbuchartikel anhand eines verweisvermittelnden Textsegmentes einen Verweis und folgt diesem dadurch, dass er eine Verweispfandhandlung erfolgreich ausführt, so dass er zum Verweiszielbereich kommt, so kann man z.B. sagen: Werner hat einen wörterbuchinternen Zugriffspfad etabliert, auf dem er vom verweisvermittelnden Textsegment zum Verweiszielbereich gelangt ist.

In diesem Beitrag wird es darum gehen, alle Typen von Zugriffspfaden unter einer einheitlichen Terminologie kennen zu lernen. Bei der Terminologiebildung greifen wir auf folgende lateinische Verben zurück: *currere* (‚laufen‘), *incurrere* (‚hineinlaufen‘), *excurrere* (‚hinauslaufen‘) und *adcurrere* (‚bis zu einem bestimmten Punkt laufen‘). Nach diesen Verben bilden wir folgende fachsprachliche Adjektive: *binnenkurrent*, *inkurrent*, *exkurrent* und *adkurrent*. Die vier Adjektive sind nicht steigerungsfähig; sie sind attributiv, prädikativ und adverbiell verwendbar. Anhand der Veranschaulichung in Abb. 1-1 lässt sich die Bedeutung dieser vier Adjektive leicht erfassen.

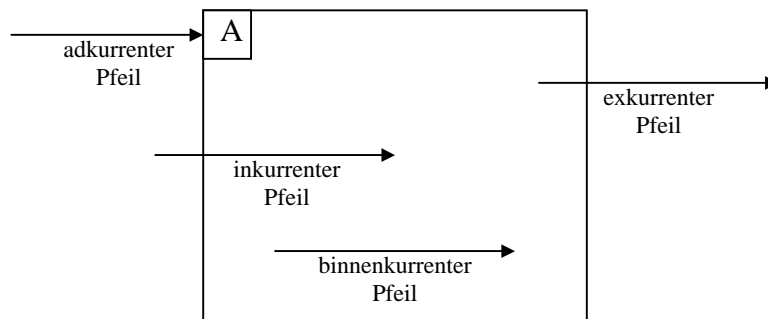


Abb. 1-1: Veranschaulichung zur Semantik der Terminologie

Ein adkurrenter Pfeil führt hin zu etwas. Ein inkurrenter Pfeil führt in etwas hinein. Ein binnenkurrenter Pfeil verläuft innerhalb von etwas, und ein exkurrenter Pfeil führt aus etwas hinaus. Es wird also jeweils nur eine Eigenschaft der Pfeile benannt, so dass offen bleibt, welche der anderen Eigenschaften ein Pfeil auch noch hat. Bevor wir die erläuterten Termini auf Zugriffspfade beziehen, werden zunächst im 2. Abschnitt einige lexikographietheoretische Begrifflichkeiten eingeführt, deren Kenntnis zum Verständnis der späteren Ausführungen im 3. Abschnitt erforderlich ist.

2. Zu einigen Schnittstellenelementen und Benutzungshandlungen

Jedes Printwörterbuch weist mindestens eine äußere Zugriffsstruktur auf. Eine äußere Zugriffsstruktur ist eine lineare Ordnungsstruktur (im mathematischen Sinn), deren Elemente äußere Zugriffstextelemente sind, auf die aufgrund der intuitiven Kenntnis der Ordnungsstrukturgesetzmäßigkeiten extern zugegriffen werden kann (vgl. Wiegand 2005). Jede äußere Zugriffsstruktur weist einen direkten Zugriffsbereich auf. In diesem liegen die direkt akzessiven lexikographischen Daten. Im Schriftkulturraum der Alphabetschriften sind die alphabetischen makrostrukturellen äußeren Zugriffsstrukturen die bekanntesten und die wichtigsten. Ihr direkter Zugriffsbereich sind Wörterverzeichnisse; ihre

äußeren Zugriffstextelemente heißen unter makro- und zugriffsstrukturellem Aspekt *Lemmata* und unter dem Aspekt, dass mit ihnen etwas angegeben wird, *lemmatische Angaben*. Zu jedem Lemma gehört ein direkter Zugriffssektor, auf den es Zugriff eröffnet. Dieser Zugriffssektor heißt *Wörterbuchartikel*. Wörterbuchartikel sind die wichtigsten lexikographischen akzessiven Einträge. Andere akzessive Einträge sind die akzessiven Außentext-, Binnentext-, Umtext-, Einschub- und die akzessiven Registereinträge. Die äußeren Zugriffstextelemente dieser akzessiven Einträge (die den Lemmata entsprechen) sind Eintragseingänge und heißen z.B. *Außentexteingang* oder *Umtexteingang*. Ein Umtexteingang ist ein Element einer äußeren Umtextzugriffsstruktur; sein Zugriffssektor ist der zugehörige akzessive Umtexteintrag. Für die Eingänge, die zu anderen Typen gehören, gilt Entsprechendes. Äußere Zugriffsstrukturen stellen die externe Datenakzessivität für diejenigen potentiellen Benutzer sicher, die die allgemeine Wörterbuchbenutzungspraxis beherrschen und daher in der Lage sind, externe Zugriffshandlungen auszuführen.

Nach der in Wiegand (1998) entworfenen Theorie der Benutzungshandlungen wird eine Benutzungshandlung h_1 , die zu Handlungstyp T_1 gehört, stets dadurch ausgeführt, dass mindestens eine andere Benutzungshandlung h_2 ausgeführt wird, die zu einem bestimmten anderen Handlungstyp T_2 gehört; man sagt dann, dass h_2 im Ausführungsbereich von h_1 liegt. Beispielsweise wird eine Nachschlagehandlung dadurch ausgeführt, dass zuerst eine externe Zugriffshandlung und dann z.B. eine interne Zugriffshandlung ausgeführt wird. Eine externe Zugriffshandlung wird dadurch ausgeführt, dass eine externe Suchhandlung ausgeführt wird; und letztere wird dadurch ausgeführt, dass der Benutzer-in-actu als Zugriffshilfe die gleichen Ordnungsgesetzlichkeiten bemüht wie der Lexikograph, der die zugriffsstrukturelle Ordnung etabliert hat. Externe Zugriffshandlungen liegen nicht nur im Ausführungsbereich von Nachschlagehandlungen, sondern auch in dem von externen Verweisbefolgungshandlungen.

Der externen Datenakzessivität steht die interne gegenüber. Erstere ist obligatorisch. Letztere ist fakultativ. Die interne Datenakzessivität wird über innere Zugriffsstrukturen hergestellt. Nur die alphabetischen inneren Zugriffsstrukturen, die sich z.B. bei den alphabetisch geordneten Literaturangaben am Ende von Fachwörterbuchartikeln finden, funktionieren auf die gleiche Weise wie externe Zugriffsstrukturen. Denn die nichtalphabetischen inneren Zugriffsstrukturen basieren nicht auf allgemeinen Ordnungsstrukturen und deren intuitiver Kenntnis, sondern auf dem Wissen, das ein kundiger Benutzer über die textuellen Strukturen der Wörterbuchartikel eines bestimmten Wörterbuchs und gegebenenfalls über die der anderen akzessiven Einträge hat.

Die wichtigsten lexikographischen Daten sind in Wörterbüchern mit kondensierten Wörterbuchartikeln (*sensu* Wiegand 2003) die artikelinternen Angaben; wir berücksichtigen im Folgenden jedoch auch Angaben in den anderen akzessiven Einträgen, wie z.B. Umtextangaben und Registerangaben (vgl. Wiegand 2005). Hinsichtlich der Angabenakzessivität lässt sich eine ganze Reihe

von Angabetypen unterscheiden. Die allermeisten Angaben gehören zum Typ der akzessiven Angabe. Eine Angabe ist dann akzessiv, wenn sie in einem der lexikographischen akzessiven Einträge steht, also z.B. in einem Wörterbuchartikel oder in einem akzessiven Umtexteintrag. Eine Angabe gehört zum Typ der nichtakzessiven Angabe, wenn sie z.B. in den Benutzungshinweisen zu Erklärungszwecken erwähnt wird. Eine Angabe gehört zum Typ der extern akzessiven Angabe, wenn sie von einem Benutzer, der von außerhalb des Wörterbuchs kommt, durch die Ausführung von externen Zugriffshandlungen erreichbar ist, die entweder im Ausführungsbereich von Nachschlagehandlungen oder in dem von Verweisbefolgungshandlungen liegen. Alle Angaben in allen Wörterbuchartikeln eines alphabetischen Wörterbuchs gehören zum Typ der extern direkt akzessiven Angabe: Auf sie kann ein Benutzer, der von außerhalb des Wörterbuchs kommt, über eine äußere Zugriffsstruktur, nämlich die alphabetische makrostrukturelle äußere Zugriffsstruktur, direkt zugreifen. Alle Lemmata als die äußeren Zugriffstextelemente und damit alle lemma-tischen Angaben gehören zum Typ der primär akzessiven Angabe: Sie werden vom Benutzer-in-actu zuerst erreicht. Alle nichtlemmatischen Angaben gehören zum Typ der sekundär akzessiven Angabe: Sie sind nur über das Lemma und im kondensierten Wörterbuchartikel nach ihm erreichbar. Entsprechendes gilt für die anderen akzessiven Einträge, die keine Wörterbuchartikel sind: Die Eintragsgänge sind primär akzessiv; die eingangsexternen Angaben sind sekundär akzessiv. Zum Typ der extern indirekt akzessiven Angabe gehören Angaben, die für einen Benutzer, der von außerhalb des Wörterbuchs kommt, nur registervermittelt erreichbar sind, weil im Register zunächst auf ihre Fundortdaten zugegriffen werden muss. Angaben, die sowohl direkt über eine äußere Zugriffsstruktur als auch registervermittelt erreichbar sind, gehören zum Typ der extern direkt und indirekt akzessiven Angabe. Den extern akzessiven stehen die extern nichtakzessiven Angaben gegenüber. Letztere finden sich in Wörterbüchern, die hinsichtlich der Angabenakzessivität Defekte aufweisen. Dies ist z.B. im Dornseiff (1934) der Fall. Dieser weist, im Unterschied zum Dornseiff (2004), kein vollständiges alphabetisches Zugriffsregister auf. Daher sind alle Angaben in den sachgruppenbezogenen Wörterbuchartikeln, die nicht als Registergänge genannt werden, extern nichtakzessive Angaben. Extern nichtakzessive Angaben können intern akzessiv sein. — Zum Typ der intern akzessiven Angabe gehören alle Angaben, die für einen Benutzer-in-actu nur dann über eine äußere Zugriffsstruktur erreichbar sind, wenn er von einem zugriffsverschiedenen Ort innerhalb eines Wörterbuchs kommt; dies ist dann der Fall, wenn er einem Verweis folgt, oder wenn ein wörterbuchinternes konfliktives Textsegment der Grund für die Benutzungshandlung ist. Ein zugriffsverschiedener Ort, an dem ein verweisvermittelndes oder ein konfliktives Textsegment begegnet, ist stets in einem anderen akzessiven Eintrag situiert. Im Terminus *intern akzessive Angabe* ist *intern* daher wie *wörterbuchintern* zu lesen. Von den wörterbuchintern akzessiven Angaben müssen die eintragsintern akzessiven Angaben, wie z.B. die artikelintern oder

umtextintern akzessiven Angaben, unbedingt unterschieden werden; denn die Letztgenannten heißen so, weil sie nicht über äußere, sondern über innere Zugriffsstrukturen erreichbar sind. Zum Typ der extern und intern akzessiven Angaben gehören alle Lemmazeichengestaltangaben, die Verweisadressenträger sind. Auf sie kann sowohl durch die Ausführung externer Zugriffshandlungen, die im Ausführungsbereich von Nachschlagehandlungen liegen, extern zugegriffen werden, als auch durch die Ausführung externer Zugriffshandlungen, die im Ausführungsbereich von Verweisbefolgungshandlungen liegen.

Angaben, die nur über eine äußere Zugriffsstruktur erreichbar sind, heißen *monoakzessive Angaben*. Sind sie von einem Benutzer-in-actu, der von außerhalb des Wörterbuchs kommt, über nur eine äußere Zugriffsstruktur erreichbar, dann liegen extern monoakzessive Angaben vor. Sind sie dagegen nur von einem Benutzer-in-actu, der von einem zugriffsverschiedenen Ort innerhalb eines Wörterbuchs kommt, über eine äußere Zugriffsstruktur erreichbar, liegen intern monoakzessive Angaben vor. Angaben, die über mehr als eine äußere Zugriffsstruktur erreichbar sind, heißen *polyakzessive Angaben*. Sind sie von einem Benutzer-in-actu, der von außerhalb des Wörterbuchs kommt, über mehrere äußere Zugriffsstrukturen erreichbar, dann liegen extern polyakzessive Angaben vor. Sind sie dagegen nur von einem Benutzer-in-actu, der von einem zugriffsverschiedenen Ort innerhalb eines Wörterbuchs kommt, über mehrere äußere Zugriffsstrukturen erreichbar, liegen intern polyakzessive Angaben vor.

Nach diesen Überlegungen kann der Typologiegraph in Abb. 2-1 präsentiert werden; da der Typologiegraph nicht kommentiert ist, sei ausdrücklich darauf hingewiesen, dass die unmittelbaren Untertypen des Typs der akzessiven Angabe nur stufen- aber nicht kriteriengleich sind.

Wir hatten bereits gesehen, dass die jeweilige Ausführung von Verweisbefolgungshandlungen in metaphorischer Sicht- und Redeweise als Bewegungen des Benutzers-in-actu innerhalb eines Wörterbuchs verstanden werden können. Damit solche Bewegungen zielgerichtet erfolgen können, weisen alle verweisvermittelnden Textsegmente, wie z.B. Verweisangaben oder Angaben mit Verweiskennzeichnung, eine mediostrukturelle Orientierung auf (vgl. Wiegand 2002); diese ist eine Eigenschaft anhand derer dem Benutzer ersichtlich ist, wo er den Verweiszielbereich findet. Sie ist dadurch gegeben, dass mit dem adressenvermittelnden Textsegment, das ein Teil des verweisvermittelnden Textsegmentes ist, die Verweisadresse genannt wird, die als mediostrukturelles Leitelement für die Ausführung einer Verweisbefolgungshandlung benötigt wird. In diesem Beitrag spielt besonders die mediostrukturelle Außenorientierung eine Rolle; diese liegt vor, wenn mit dem adressenvermittelnden Textsegment eine Verweisaußenadresse genannt wird. Verweisaußenadressen sind solche Verweisadressen, die nicht im gleichen akzessiven Eintrag zu finden sind, in dem die verweisvermittelnden Textsegmente stehen, mit dem sie genannt werden. Wird beispielsweise mit einer Verweisadressenangabe einer artikelinternen Verweisangabe eine Umtextaußenadresse genannt, die in einer

äußeren Umtextzugriffsstruktur liegt, dann ist die gegebene mediostrukturelle Außenorientierung eine Umtextorientierung, und es liegt eine umtextorientierte Verweisangabe vor. Entsprechend gibt es u.a. artikelinterne binnentextorientierte, artikelorientierte und einschuborientierte Verweisangaben. Erschließt ein Benutzer-in-actu beispielsweise anhand einer umtextorientierten Verweisangabe einen Verweis, dann ist dieser ebenfalls umtextorientiert. Folgt er dem umtextorientierten Verweis dadurch, dass er eine externe Verweisbefolgungshandlung ausführt und ist diese erfolgreich, so dass er die Verweisadresse gefunden hat, dann hat der erfolgreiche Benutzer einen Zugriffspfad etabliert, der ebenfalls umtextorientiert ist. — Damit sind wir bei den Zugriffspfaden, die im nächsten Abschnitt z.T. anhand von Beispielen systematisch betrachtet werden.

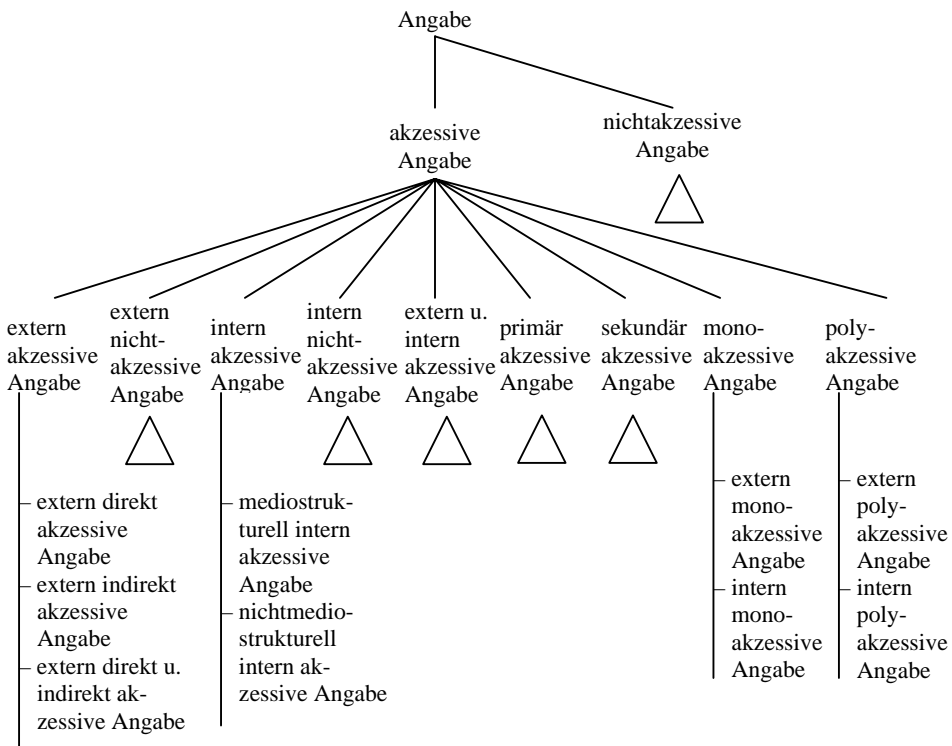


Abb. 2-1: Nichtkommentierter Typologigraph zu einem Ausschnitt aus einer Typologie von Angaben nach der externen Angabenakzessivität

3. Zugriffspfade in Printwörterbüchern

Die wichtigste Voraussetzung dafür, dass Zugriffspfade von Benutzern-in-actu im Zuge der Ausführung von Zugriffshandlungen etabliert werden können, ist

die Existenz von Zugriffsstrukturen; denn nur sie erlauben einen Zugriff auf lexikographische Daten, mit dem eine erfolgreich ausgeführte Zugriffshandlung abgeschlossen wird. Ein Zugriff ist das Ergebnis einer systematischen Suche bei der Ausführung von Zugriffshandlungen unter Zuhilfenahme von Ordnungsgesetzlichkeiten und von Wissen über die textuellen Strukturen in akzessiven Einträgen. Bei den Zugriffspfaden werden die einfachen Zugriffspfade von den zusammengesetzten unterschieden. Für Erstere kann folgende Definition (= D) gegeben werden:

(D 3-1: einfacher Zugriffspfad)

Ein *einfacher Zugriffspfad* ist ein von einem Benutzer-in-actu im Zuge der Ausführung einer erfolgreichen Zugriffshandlung etablierter Pfad, für dessen Pfadanfang und Pfadende folgende Bedingungen gelten:

- (a) Ist der Pfadanfang wörterbuchextern, dann ist das Pfadende wörterbuchintern;
- (b) Ist der Pfadanfang wörterbuchintern, dann ist das Pfadende entweder wörterbuchintern oder -extern.

Der Terminus *einfacher Zugriffspfad* wird oberbegrifflich verwendet, und zwar für einfache äußere und einfache innerer Zugriffspfade. Zusammengesetzte Zugriffspfade gibt es nur bei den äußeren Zugriffspfaden.

3.1 Äußere Zugriffspfade

Äußere Zugriffspfade haben ihren Namen daher, dass sie zu und durch Abschnitte von *äußeren* Zugriffsstrukturen führen und nicht etwa daher, dass ein Benutzer von außerhalb des Wörterbuchs kommt. Es können einfache äußere von zusammengesetzten äußeren Zugriffspfaden unterschieden werden; wir betrachten zunächst die Ersteren. Auf die Letzteren wird in 3.1.2 eingegangen.

3.1.1 Einfache äußere Zugriffspfade

Einfache äußere Zugriffspfade sind diejenigen Zugriffspfade, die bei der Wörterbuchbenutzung am häufigsten etabliert werden. Ein einfacher äußerer Zugriffspfad kann wie folgt definiert werden:

(D 3-2: einfacher äußerer Zugriffspfad)

Ein *einfacher äußerer Zugriffspfad* ist ein benutzerspezifischer Weg, der durch eine Reihe von äußeren Zugriffstextelementen einer äußeren Zugriffsstruktur zu einem gesuchten äußeren Zugriffstextelement führt und der durch die erfolgreiche Ausführung einer externen Zugriffshandlung etabliert wird.

Ein einfacher äußerer Zugriffspfad hat stets zwei Pfadstationen: einen Pfadanfang und ein Pfadende. Ein Pfadanfang ist entweder der wörterbuchexterne

textuelle Ort — wie im Beispiel B 3-1 — oder derjenige kognitive Zustand, der als Basis für die Erfassung des Leitelements dient, das ein Benutzer für die Ausführung einer externen Zugriffshandlung benötigt, die im Ausführungsbereich einer Nachschlagehandlung liegt; oder der Pfadanfang eines einfachen äußeren Zugriffspfades ist wörterbuchintern situiert. Dabei muss zwischen zwei Fällen unterschieden werden; der erste ist gegeben, wenn — wie in den Beispielen B 3-2 bis B 3-4 — der Pfadanfang ein verweisvermittelndes Textsegment ist, dessen adressenvermittelndes textuelles Teilsegment die Verweisadresse liefert, die als mediostrukturelles Leitelement dient. Der zweite Fall liegt vor, wenn — wie im Beispiel B 3-5 — der Pfadanfang ein konfliktives Textsegment ist, welches das Leitelement für die Ausführung einer Konsultationshandlung liefert, in deren Ausführungsbereich eine externe Zugriffshandlung liegt.

Nach diesen Erörterungen lassen sich zunächst die nachfolgend genannten Typen von einfachen äußeren Zugriffspfaden unterscheiden, wobei bei der Bildung der Termini für die Zugriffspfadtypen die in Abb. 1-1 erläuterten Adjektive verwendet werden.

- (1) Der Typ des adkurrenten äußeren Zugriffspfades: Zu diesem Zugriffspfadtyp gehören alle einfachen Zugriffspfade, deren Pfadanfang wörterbuchextern und deren Pfadende ein äußeres Zugriffstextelement ist, so dass es wörterbuchintern situiert ist. Dieser Zugriffspfadtyp gehört zu den nichtmediostrukturellen Zugriffspfadtypen (vgl. 3.1.1.3).
- (2) Der Typ des binnenkurrenten äußeren Zugriffspfades: Zu diesem Zugriffspfadtyp gehören alle einfachen Zugriffspfade, deren Pfadanfang und Pfadende wörterbuchintern situiert ist. Zum zweiten Typ gehören zwei Untertypen:
 - (2.1) Der Typ des mediostrukturellen binnenkurrenten äußeren Zugriffspfades: Zu diesem Zugriffspfadtyp gehören alle einfachen Zugriffspfade, deren Pfadanfang ein verweisvermittelndes Textsegment und deren Pfadende wörterbuchintern situiert ist.
 - (2.2) Der Typ des nichtmediostrukturellen binnenkurrenten äußeren Zugriffspfades: Zu diesem Zugriffspfadtyp gehören alle einfachen Zugriffspfade, deren Pfadanfang ein wörterbuchinternes konfliktives Textsegment und dessen Pfadende wörterbuchintern situiert ist.

In den folgenden Abschnitten werden zu den Typen (1) bis (2.2) Beispiele betrachtet. Dabei ist zu beachten, dass die vier bisher eingeführten Typen in einer Typologie der Zugriffspfade auf den obersten Typologiestufen stehen, so dass die Beispiele zugleich Beispiele für einen von deren Untertypen sind. In den Beispielüberschriften stehen die Termini für diese Untertypen hinter dem Schrägstrich.

3.1.1.1 Adkurrente äußere Zugriffspfade

Systematisch betrachtet, gehören alle adkurrenten äußeren Zugriffspfade zu den nichtmediostrukturellen Zugriffspfaden, die später in Abschnitt 3.1.1.3 behandelt werden. Ihre Betrachtung wurde jedoch vorgezogen: Nach meiner Einschätzung erleichtert ihre frühzeitige Erklärung das Verständnis der späteren Ausführungen.

Gegeben sei folgendes Beispiel:

(B 3-1: Adkurrenter äußerer Zugriffspfad / artikeladkurrenter äußerer Zugriffspfad)

Werner liest einen Text und stößt dort auf das Wort *karamell*. Mit der Suchfrage „Schreibt man das Adjektiv tatsächlich mit zwei l?“ greift Werner zum Duden-²GW und findet dort den Artikel:

wa1: **Ka|ra|me|ll**, <indekl. Adj.>: *bräunlich gelb*.

Um das Lemma zu finden, hat Werner eine externe Zugriffshandlung mit dem Leitelement [karamell] ausgeführt, die erfolgreich war. Im Zuge der Ausführung dieser Benutzungshandlung hat er einen Zugriffspfad etabliert, der zu einem Untertyp des Typs des adkurrenten äußeren Zugriffspfades gehört, nämlich zum Typ des artikeladkurrenten äußeren Zugriffspfades. Zu diesem Untertyp gibt es auf der gleichen Typologiestufe mehrere kriteriengleiche Untertypen. Der entsprechende Ausschnitt aus einer Typologie der Zugriffspfade findet sich in Abb. 3-1.

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3-1 gehören, seien die beiden Folgenden angegeben:

(D 3-3: adkurrenter äußerer Zugriffspfad)

Ein *adkurrenter äußerer Zugriffspfad* ist ein nichtmediostruktureller einfacher äußerer Zugriffspfad, dessen Pfadanfang der jeweilige wörterbuchexterne textuelle Ort oder der jeweilige kognitive Zustand ist, der die Basis für die Erfassung des Leitelementes zur Ausführung einer externen Zugriffshandlung bildet und dessen Pfadende ein äußeres Zugriffstextelement ist.

(D 3-4: umtextadkurrenter äußerer Zugriffspfad)

Ein *umtextadkurrenter äußerer Zugriffspfad* ist ein nichtmediostruktureller einfacher äußerer Zugriffspfad, dessen Pfadanfang der jeweilige wörterbuchexterne textuelle Ort oder der jeweilige kognitive Zustand ist, der die Basis für die Erfassung des Leitelementes zur Ausführung einer externen Zugriffshandlung bildet, und dessen Pfadende ein Umtexteingang ist.

Der genaue Verlauf eines bestimmten konkreten Zugriffspfades, z.B. der des artikeladkurrenten äußeren Zugriffspfades, den Werner in B 3-1 etabliert hat,

ist empirisch schwer festzustellen. Die beste Methode hierzu ist die der periactionalen Benutzungskommentierung (*sensu* Wiegand 1998: 1015ff). Wendet man diese so an, dass man Testpersonen gleiche Zugriffsaufgaben stellt, dann zeigt sich, dass die konkreten Zugriffspfade im Detail sehr verschieden sind und nur darin übereinstimmen, dass der Pfadanzug und das Pfadende gleich sind. Es hat sich auch gezeigt, dass nur erfahrene Benutzer die kürzesten Zugriffspfade wählen, so dass sich möglichst kurze äußere Zugriffszeiten ergeben.

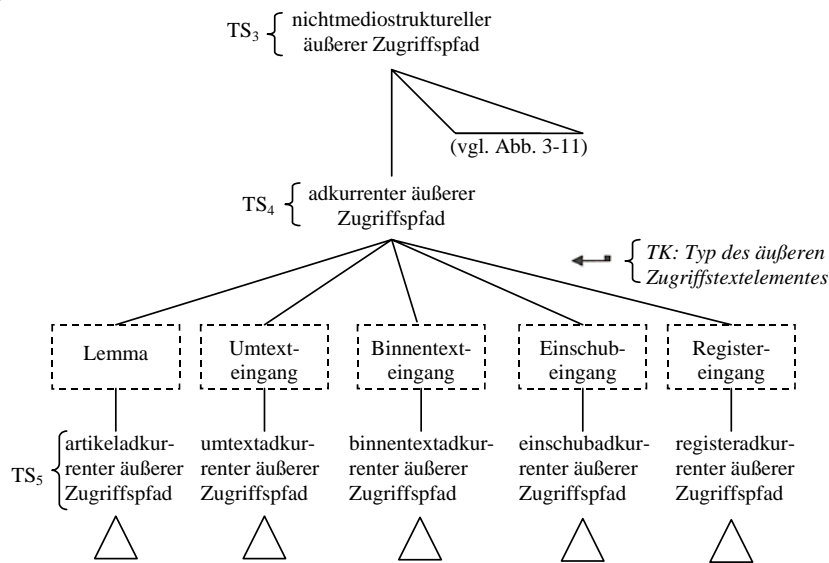


Abb. 3-1: Erster partiell erweitert kommentierter Typologieglyph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden; *Abkürzungen:* TS = Typologiestufe; TK = Typologiekriterium; „← TK “ bedeutet soviel wie *die Anwendung des TK führt zu der Unterteilung*

Zu jedem Zugriffspfadtyp lässt sich ein allgemeines Zugriffspfadbild angeben. Während alle allgemeinen Artikeltextstrukturbilder, Mikrostrukturbilder und Suchbereichsstrukturbilder formale Darstellungsmittel sind, die nach festen Konventionen gebildete Ausschnitte aus geordneten und kommentierten Baumgraphen darstellen, sind allgemeine Zugriffspfadbilder Veranschaulichungen, die allerdings ebenfalls nach festgelegten Darstellungskonventionen hergestellt sind (vgl. Wiegand 2007). Das allgemeine Strukturbild für den Typ des artikeladkurrenten äußeren Zugriffspfad findet sich in Abb. 3-2.

Wie die Abb. 3-2 zeigt, werden in einem allgemeinen Zugriffspfadbild der Pfadanzug, der Pfadverlauf durch einen Pfeil, das Pfadende sowie der durchlaufene Ausschnitt der zugehörigen äußeren Zugriffsstruktur repräsentiert, und die Bildelemente werden mit ihren Termini versehen.

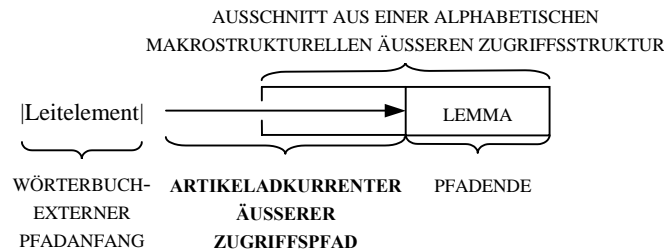


Abb. 3-2: Allgemeines Zugriffspfadbild für den Typ des artikeladkurrenten äußeren Zugriffspfades

3.1.1.2 Mediostrukturelle äußere Zugriffspfade

Bei den mediostrukturellen äußeren Zugriffspfaden werden hier nur die binnenkurrenten von den wörterbuchekurrenten unterschieden. Weitere Unterscheidungen sind möglich, wenn man mehr als ein Wörterbuch zusammen betrachtet.

3.1.1.2.1 Mediostrukturelle binnenkurrente äußere Zugriffspfade

Bei den mediostrukturellen binnenkurrenten äußeren Zugriffspfaden ist die typologische Vielfalt am größten. Dies hängt damit zusammen, dass im Unterschied z.B. zu den adkurrenten äußeren Zugriffspfaden die wörterbuchinterne textuelle Lage des Pfadanfanges (woher der Pfad kommt), die mediostrukturelle Orientierung des Pfades und die textuelle Lage des Pfadendes (wohin der Pfad führt) bei der Typologiebildung Berücksichtigung finden.

Gegeben sei das folgende Beispiel:

(B 3-2: Binnentextekurrenter Zugriffspfad/binnentextekurrenter artikeladkurrent orientierter Zugriffspfad)

Werner studiert den eingelagerten Binnentext mit dem Binnentexttitel „Fach- und Sondersprachen“ in ²Herders Sprachbuch. Ein Abschnitt des Binnentextes endet wie folgt:

„Andererseits sind Wörter aus dem technischen Bereich, die ursprünglich berufssprachliche Prägungen waren, in den allgemeinen Wortschatz eingegangen, weil die Dinge und Verfahren, die sie bezeichnen, den meisten Zeitgenossen bekannt geworden sind (↗ lüten, ↗ schweißen, ↗ fräsen, ↗ Zarge, ↗ Kardanwelle usw. □ Wortschatz, □ Wortbildung).“

Werner weiß nicht, was eine Zarge ist. Anhand der artikelorientierten Verweisangabe „↗ Zarge“ erschließt er einen artikelorientierten Verweis. Diesem folgt er dadurch, dass er unter Zuhilfenahme der Verweisadresse |Zarge| als mediostrukturellem Leitelement eine externe Verweisbefolgungshandlung dadurch

ausführt, dass er eine externe Zugriffshandlung ausführt. Diese beendet er erfolgreich mit dem Zugriff auf das Lemma „Zar|ge“, das als Verweisadressenträger fungiert. Damit hat Werner einen Zugriffspfad etabliert, der zum Typ des binnentextkurrenten artikeladkurrent orientierten Zugriffspfades gehört: Dieser ist einer der Untertypen (2. Stufe) des Typs des binnentextkurrenten Zugriffspfades. — Im Folgenden betrachten wir ein weiteres Beispiel für einen mediostrukturellen binnenkurrenten äußeren Zugriffspfad.

(B 3-3: *Artikelexkurrenter Zugriffspfad/artikelexkurrenter umtextadkurrent orientierter Zugriffspfad*)

Werner konsultiert folgenden Artikel aus dem DCE:

wa2: **darts** /da:ts || darts/ n [U] any of several games in which DARTS¹ (1) are thrown at a DARTBOARD

Anhand der umtextorientierten reduzierten Verweisangabe „U“ erschließt er einen umtextorientierten Verweis; unter Zuhilfenahme der Umtextaußenadresse [U] als mediostrukturellem Leitelement führt er eine externe Verweisbefolgungshandlung dadurch aus, dass er eine externe Zugriffshandlung ausführt, die mit dem Zugriff auf den Umtexteingang „U“ in der äußeren Umtextzugriffsstruktur erfolgreich endet. Damit hat Werner einen artikelexkurrenten umtextadkurrent orientierten Zugriffspfad etabliert. Das allgemeine Zugriffspfadbild für diesen Pfadtyp findet sich in Abb. 3-3.

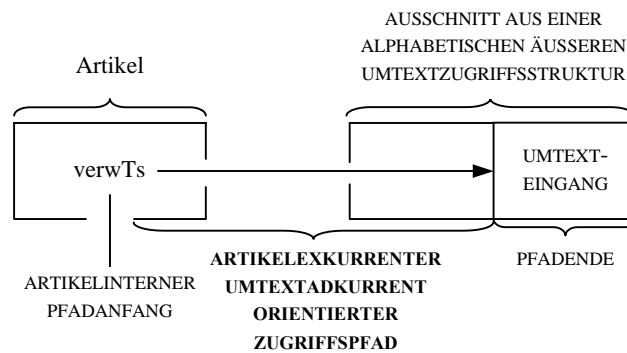


Abb. 3-3: Allgemeines Zugriffspfadbild für den Typ des artikelexkurrenten umtextadkurrent orientierten Zugriffspfades; *Abkürzungen:* verwTs = verweisvermittelndes Textsegment

Ein drittes Beispiel für einen mediostrukturellen binnenkurrenten äußeren Zugriffspfad ist das Folgende:

(B 3-4: *Artikelexkurrenter Zugriffspfad/artikelexkurrenter artikelinkurrent orientierter Zugriffspfad*)

Werner beschäftigt sich mit Diminution. In dem Wörterbuchartikel zum Lemmazeichen *Tierchen* im Duden-GW findet er folgende usuelle vollständige Verweisangabe (a) „↑Tier (1)“. Die abstrakte hierarchische und isomorphe konkrete hierarchische Angabemikrostruktur der Verweisangabe (a) ist in Abb. 3-4 dargestellt.

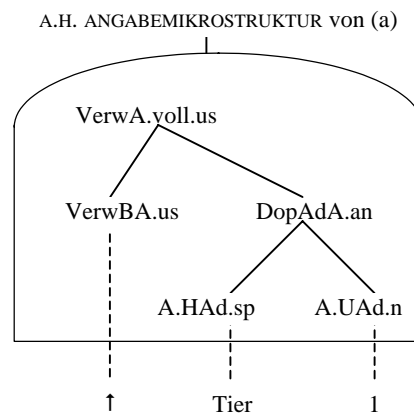


Abb. 3-4: Kommentierter Strukturgraph zur abstrakten hierarchischen (und isomorphen konkreten hierarchischen) Angabemikrostruktur der Verweisangabe (a). *Abkürzungen:* A. = abstrakt; H. = hierarchisch; VerwA.voll.us = usuelle vollständige Verweisangabe; VerwBA.us. = usuelle Verweisbeziehungsangabe; DopAdA.an = alphanumerische Doppeladressenangabe; A.HAd.sp = sprachliche Hauptadressenangabe; A.UAd.n = numerische Unteradressenangabe; „X—Y“ bedeutet (von unten nach oben gelesen) so viel wie X ist eine Teilangabe von Y; „U V“ bedeutet so viel wie U ∈ V (U ist ein Element von V)

Wird anhand von verweisvermittelnden Textsegmenten, mit denen eine Doppeladresse genannt wird, ein Verweis erschlossen und schließt sich eine erfolgreiche Verweisbefolgungshandlung an, wird stets ein mediostruktureller Zugriffspfad etabliert, der eine inkurrente mediostrukturelle Orientierung aufweist, so dass er in einen akzessiven Eintrag oder in einem anderen Zugriffssektor (vgl. 3.3) hineinführt. Die inkurrent orientierten Zugriffspfade werden — obwohl ihr Pfadende innerhalb akzessiver Einträge liegt — zu den äußeren Zugriffspfaden gerechnet, da sie stets über äußere Zugriffstextelemente führen.

In B 3-4 erschließt Werner anhand der artikelorientierten Verweisangabe einen artikelorientierten Verweis. Diesem folgt er dadurch, dass er unter Zuhilfenahme der Doppeladresse |Tier 1| als mediostrukturellem Leitelement eine externe Verweisbefolgungshandlung dadurch ausführt, dass er eine externe Zugriffshandlung ausführt. Diese ist erfolgreich, und Werner kommt zunächst zum Lemma des folgenden Wörterbuchartikels:

wa₃: **Tier** [...], das; -[e]s, -e [...] **1.** mit Sinnes- und Atmungsorganen ausge-

*stattetes, sich von anderen tierischen od. pflanzlichen Organismen ernähren-
endes, in der Regel frei bewegliches Lebewesen [...]*

Nach Erreichung des Lemmas „Tier“ führt Werner unter Zuhilfenahme der Unteradresse |1| eine interne Zugriffshandlung aus, die ihn zum Pfadende, der Polysemieangabe „1.“, bringt. Damit hat Werner einen artikelexkurrenten artikelinkurrent orientierten Zugriffspfad etabliert. Das allgemeine Zugriffspfadbild für den zugehörigen Zugriffspfadtyp findet sich in Abb. 3-5.

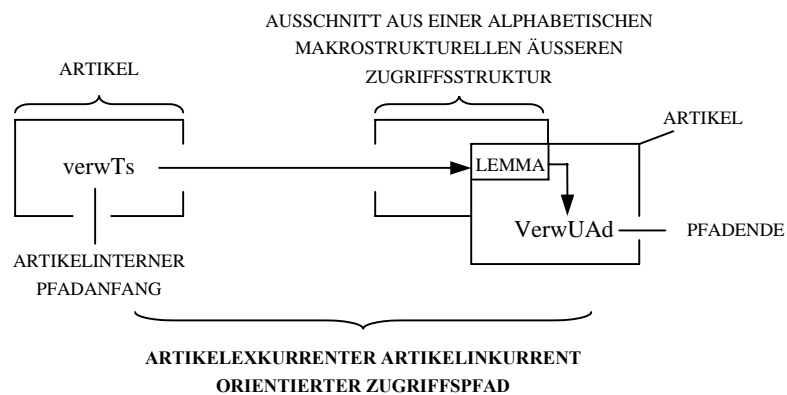


Abb. 3-5: Allgemeines Zugriffspfadbild für den Typ des artikelexkurrenten artikelinkurrent orientierten Zugriffspfades; *Abkürzungen:* verwTs = verweisvermittelndes Textsegment; VerwUAd = Verweisunteradresse

Um über die bereits erwähnte typologische Vielfalt, die bei den mediostrukturellen binnenkurrenten äußeren Zugriffspfaden gegeben ist, einen Überblick zu gewinnen, werden im Folgenden die wichtigsten Typologieausschnitte betrachtet. Mit dem Typologiekriterium „Lage des außenorientierten Pfadanfangs“ bilden wir zunächst einen Typologieausschnitt unter dem Aspekt, zu welchem Typ von akzessiven Einträgen welcher Typ von Zugriffspfaden gehört ohne Rücksicht auf die mediostrukturelle Orientierung; dieser findet sich — eingebettet in seine typologische Umgebung — in Abb. 3-6.

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3-6 gehören, seien die beiden folgenden genannt:

(D 3-5: exkurrenter Zugriffspfad)

Ein *exkurrenter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadanfang in einem akzessiven Eintrag oder einer Textverbundkonstituente liegt, in dem sein wörterbuchinternes Pfadende nicht situiert ist.

(D 3-6: *artikelexkurrenter Zugriffspfad*)

Ein *artikelexkurrenter Zugriffspfad* ist ein *mediostruktureller binnenkurrenter äußerer Zugriffspfad*, dessen Pfadende nicht im gleichen Wörterbuchartikel liegt wie sein Pfadanzfang.

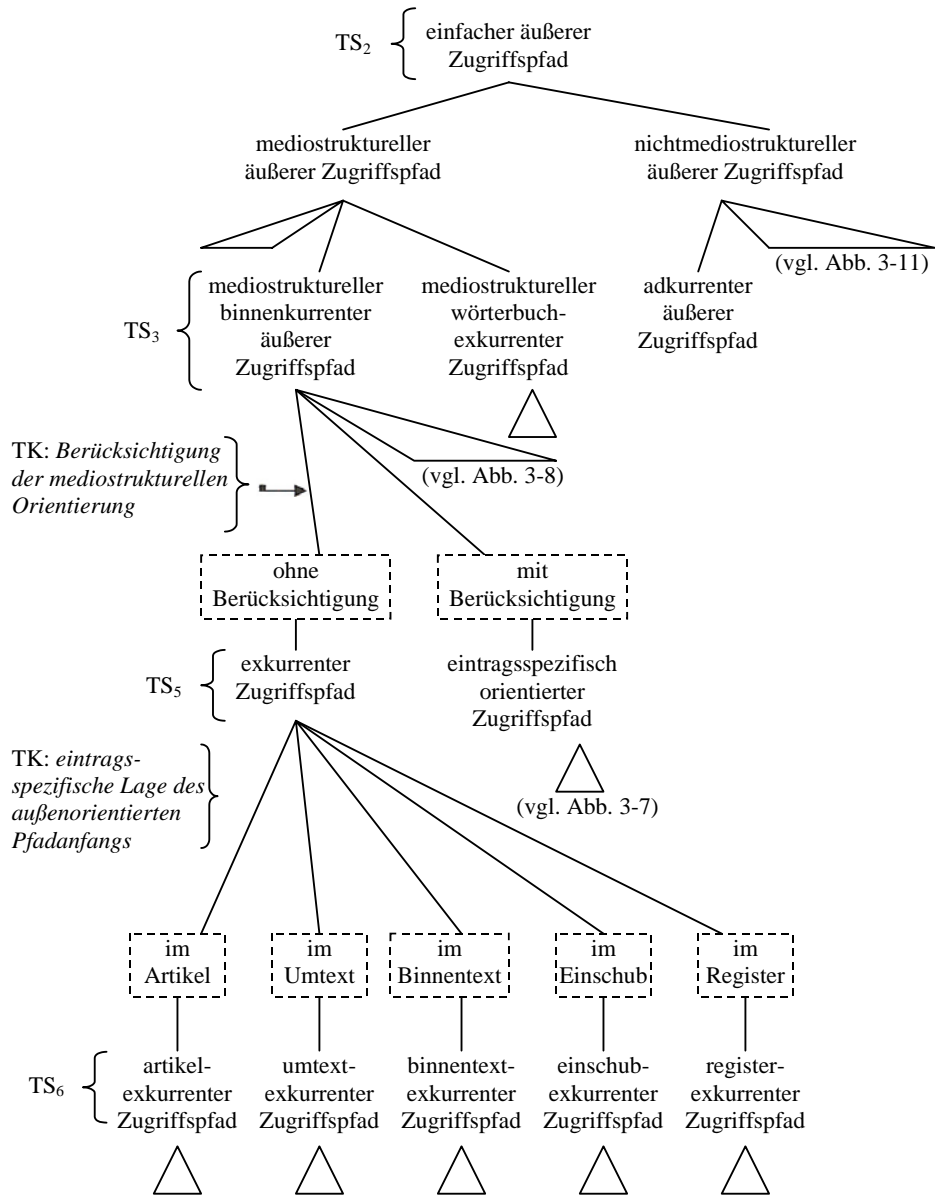


Abb. 3-6: Zweiter partiell erweitert kommentierter Typologieglyph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden

Die Frage, wozu ein Zugriffspfadtyp nützlich ist, bei dessen Bildung die mediostrukturelle Orientierung nicht berücksichtigt wird, ist berechtigt; sie wird aber wahrscheinlich nur von jemand gestellt, der sich noch nicht vor die Aufgabe gestellt sah, das qualitativ-quantitative mediostrukturelle Profil z.B. eines Lern- oder Konsultationswörterbuchs (*sensu* Wiegand 2006; vgl. Schierholz und Wiegand 2005) zu berechnen. Dann muss er nämlich z.B. angeben können, wie viele pfadanfangsverschiedene exkurrente Zugriffspfadtypen gegeben sind, wie viel Prozent davon z.B. artikelexkurrent und wie viel Prozent davon umtextexkurrent sind.

Beim nächsten Typologieausschnitt in Abb. 3-7 wird die mediostrukturelle Orientierung der mediostrukturellen binnenkurrenten äußeren Zugriffspfade berücksichtigt. In den Termini für Zugriffspfadtypen tritt dann stets eine Form von *orientiert* auf.

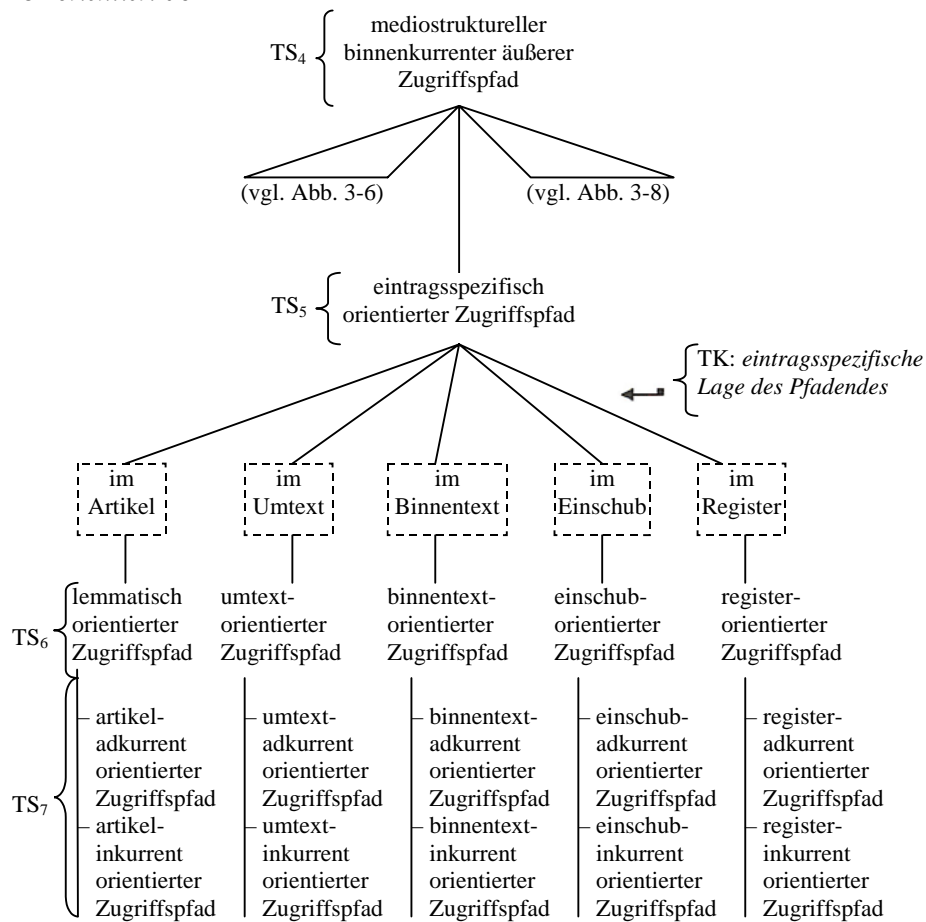


Abb. 3-7: Dritter partiell erweitert kommentierter Typologiegraph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3-7 gehören, sei die Folgende genannt:

(D 3-7: umtextorientierter Zugriffspfad)

Ein *umtextorientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadende in einem Umtext liegt.

Es ist klar, dass für alle Zugriffspfadtypen, deren Pfadende ein Element einer äußeren Zugriffsstruktur ist, ein gemeinsamer Obertyp gebildet werden kann, nämlich der Typ des adkurrent orientierten Zugriffspfades, der wie folgt definiert ist:

(D 3-8: adkurrent orientierter Zugriffspfad)

Ein *adkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadende ein Element einer äußeren Zugriffsstruktur ist.

Weiterhin ist wohl deutlich, dass auch für alle Zugriffspfadtypen, deren Pfadende nicht in einer äußeren Zugriffsstruktur liegt, aber nur über ein äußeres Zugriffstextelement erreichbar ist, ein Obertyp gebildet werden kann, nämlich der Typ des inkurrent orientierten Zugriffspfades, der wie folgt definiert ist:

(D 3-9: inkurrent orientierter Zugriffspfad)

Ein *inkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadende innerhalb eines akzessiven Eintrages liegt und nur über dessen äußeres Zugriffstextelement erreichbar ist.

Entsprechend ergibt sich der Typologieausschnitt in Abb. 3-8.

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3-8 gehören, seien die beiden folgenden genannt:

(D 3-10: binnentextadkurrent orientierter Zugriffspfad)

Ein *binnentextadkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadende ein Binnentexteingang eines akzessiven Binnentexteintrags ist.

(D 3-11: umtextinkurrent orientierter Zugriffspfad)

Ein *umtextinkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadende in einem akzessiven Umtexteintrag liegt und nur über den zugehörigen Umtexteingang erreichbar ist.

Will man bei der Bildung von mediostrukturellen Zugriffspfadtypen die Lage des Pfadanfangs, die mediostrukturelle Orientierung und die Lage des Pfad-

endes zusammen berücksichtigen, lässt sich dies am einfachsten dadurch bewerkstelligen, dass man den Typologieausschnitt in Abb. 3-6 um weitere Untertypen erweitert, so dass sich der Typologiegraph in Abb. 3-9 auf der nächsten Seite ergibt.

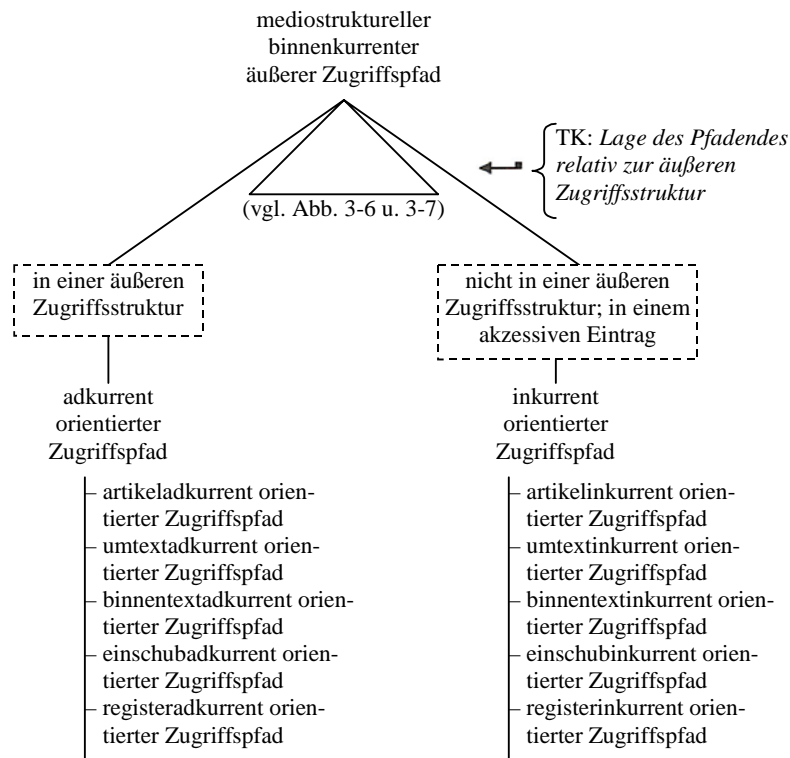


Abb. 3-8: Vierter partiell erweitert kommentierter Typologiegraph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden

Von den Definitionen, die zu dem Typologieausschnitt in Abb. 3-9 gehören, seien die beiden folgenden genannt:

(D 3-12: umtextekurrenter artikeladkurrent orientierter Zugriffspfad)

Ein *umtextekurrenter artikeladkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadanfang in einem Umtext liegt und dessen Pfadende ein Lemma ist.

(D 3-13: registerekurrenter umtextinkurrent orientierter Zugriffspfad)

Ein *registerekurrenter umtextinkurrent orientierter Zugriffspfad* ist ein mediostruktureller binnenkurrenter äußerer Zugriffspfad, dessen Pfadanfang in einem Register liegt und dessen umtextinternes Pfadende nur über einen Umtexteingang erreichbar ist.

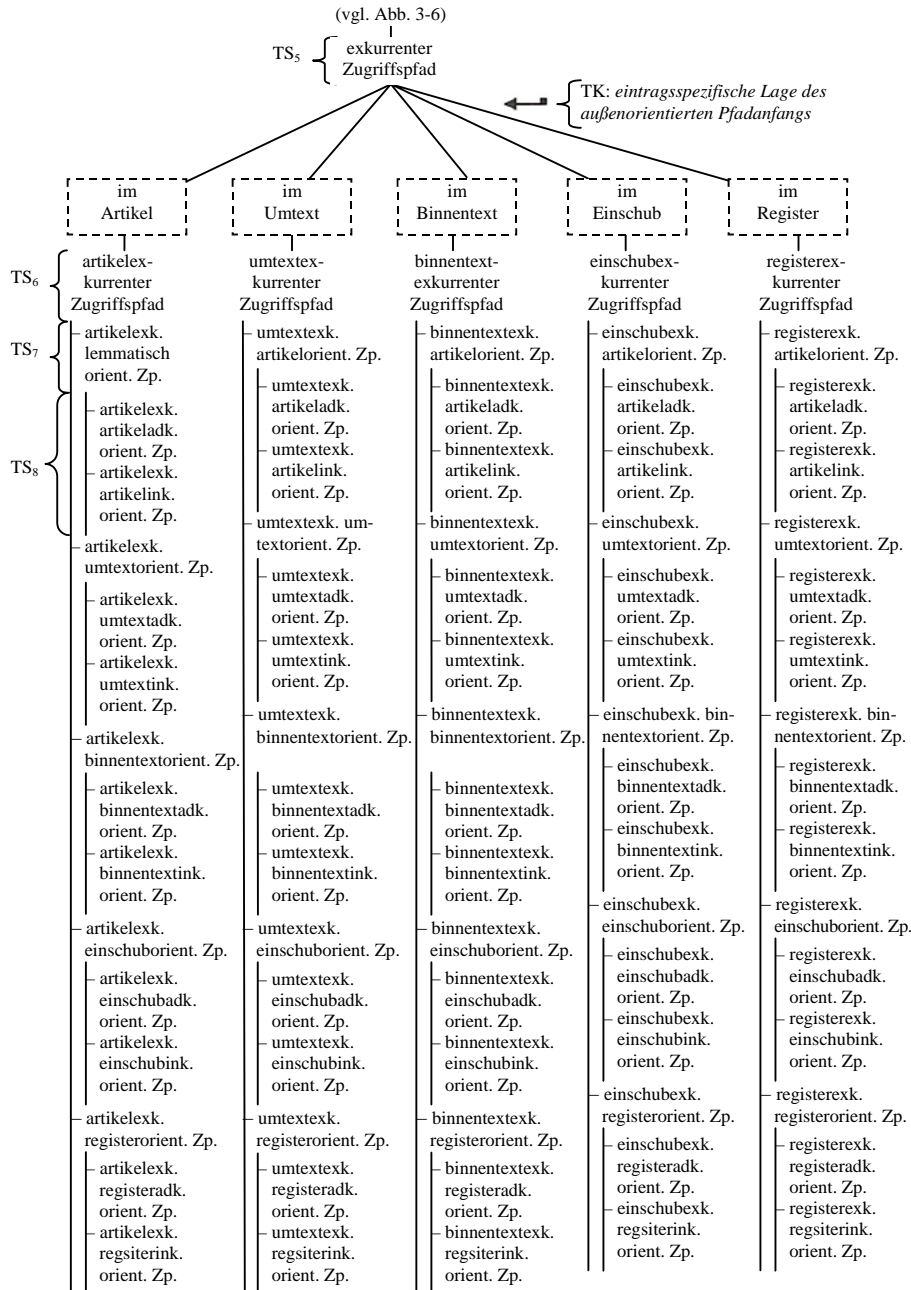


Abb. 3-9: Fünfter partiell erweitert kommentierter Typologiegraph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden; *Abkürzungen*: exk. = exkurrent; adk. = adkurrent; ink. = inkurrent; Zp. = Zugriffspfad

3.1.1.2.2 Mediostrukturelle wörterbuchekurrente äußere Zugriffspfade

Auf den Typ des mediostrukturellen wörterbuchexternen äußeren Zugriffspfades kann hier (aus Platzgründen) nur sehr kurz eingegangen werden. Auch auf die Präsentation eines Typologieausschnittes muss verzichtet werden. Mit dem Pfadanfang mediostruktureller wörterbuchekurrenter Zugriffspfade wird immer eine wörterbuchexterne Verweisadresse genannt. Solche Verweisadressen sind z.B. Quellentitel, Titel von fachlichen Publikationen oder Belegstellen. Die mit dem jeweiligen verweisvermittelnden Textsegment genannten wörterbuchexternen Adressen sind meistens verdichtet; dann müssen sie in Umtexten, wie z.B. Quellenverzeichnissen oder Literaturverzeichnissen, ergänzt werden, bevor Verweisbefolgungshandlungen ausgeführt werden. Für historische Wörterbücher beispielsweise gilt, dass sie Schnittstellenelemente aufweisen, die die Benutzer in die Lage versetzen, mediostrukturelle wörterbuchekurrente quellenorientierte Zugriffspfade zu etablieren. Für Fachwörterbücher gilt beispielsweise, dass sie Schnittstellenelemente aufweisen sollten, um wörterbuchekurrente fachliteraturorientierte Zugriffspfade zu etablieren.

3.1.1.3 Nichtmediostrukturelle äußerer Zugriffspfade

Bei den nichtmediostrukturellen äußeren Zugriffspfaden werden — neben den bereits in Abschnitt 3.1.1.1 behandelten adkurrenten äußeren Zugriffspfaden — die nichtmediostrukturellen binnenkurrenten äußeren Zugriffspfade und die nichtmediostrukturellen wörterbuchekurrenten Zugriffspfade unterschieden.

3.1.1.3.1 Nichtmediostrukturelle binnenkurrente äußere Zugriffspfade

Zugriffspfade, die zum Typ des nichtmediostrukturellen binnenkurrenten äußeren Zugriffspfades gehören, werden dann von Benutzern etabliert, wenn es bei der Wörterbuchbenutzung zu Textrezeptionsstörungen kommt und der Benutzer diese dadurch beheben möchte, dass er das gleiche Wörterbuch benutzt, anhand dessen die Textrezeptionsstörung entstanden ist, wie in folgendem Beispiel:

(B 3-5: Nichtmediostruktureller binnenkurrenter äußerer Zugriffspfad/artikel-exkurrenter und artikeladkurrenter äußerer Zugriffspfad)

Werner liest einen englischen Text und kennt das Wort *howl* nicht. Er schlägt im WNED nach und findet im Artikel mit dem Lemma „**howl**“ folgende Bedeutungsparaphrasenangabe: „to make a loud sustained doleful sound“. In dieser Angabe kennt er „sustained“ nicht. Werner führt daraufhin mit dem Leitelement |sustain| eine externe Zugriffshandlung erfolgreich aus, so dass er im WNED zum Lemma „**sus-tain**“ kommt. Damit hat er einen artikellexkurrenten und artikeladkurrenten Zugriffspfad etabliert. Das zugehörige allgemeine Zugriffspfadbild findet sich in Abb. 3-10.

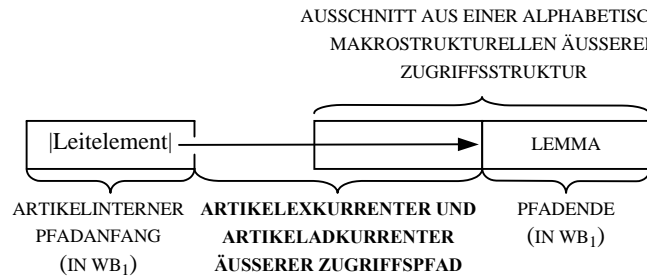


Abb. 3-10: Allgemeines Zugriffspfadbild für den Typ des artikelexkurrenten und artikeladkurrenten äußeren Zugriffspfades; *Abkürzungen:* WB = Wörterbuch

Einige wichtige Untertypen des Typs des nichtmediostrukturellen binnenkurrenten äußeren Zugriffspfades sind die in Abb. 3-11 zu einem Typologieausschnitt zusammengefasst.

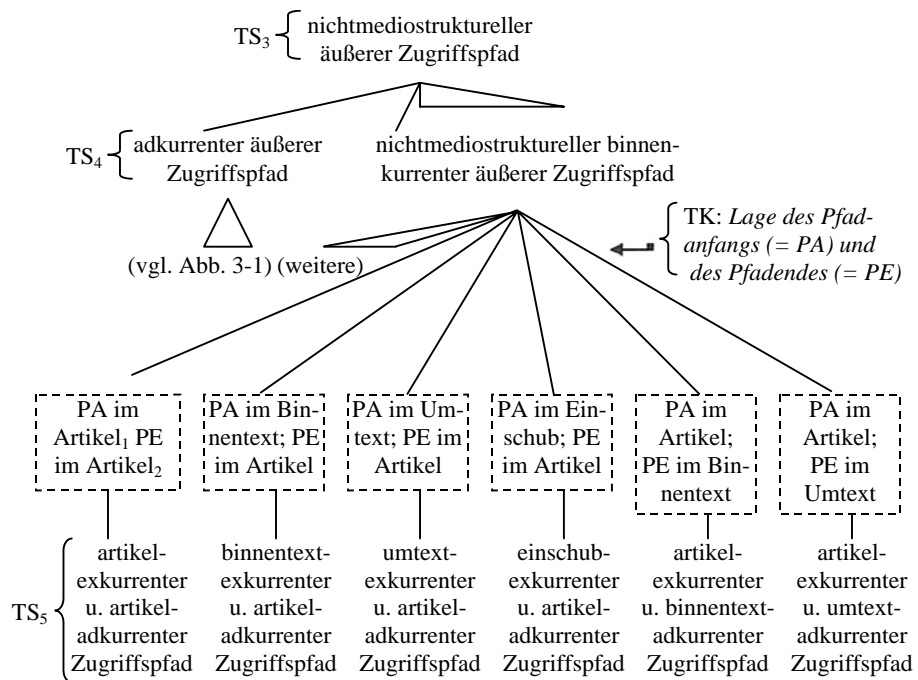


Abb. 3-11: Sechster partiell erweitert kommentierter Typologiegraph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden

3.1.1.3.2 Nichtmediostrukturelle wörterbuchekurrente äußere Zugriffspfade

Zugriffspfade, die zum Typ des nichtmediostrukturellen wörterbuchekurrenten äußeren Zugriffspfadestypes gehören, werden dann von Benutzern etabliert, wenn es bei der Benutzung eines Wörterbuchs Wb_1 zu den Textrezeptionsstörungen kommt und der Benutzer diese dadurch beheben möchte, dass er ein anderes Wörterbuch Wb_2 benutzt. Solche Situationen treten vor allem bei der Benutzung von zweisprachigen Wörterbüchern sowie bei der Benutzung von Fachwörterbüchern auf. Die nichtmediostrukturellen wörterbuchekurrenten äußeren Zugriffspfade, die öfter etabliert werden, gehören entweder zu den artikel- und wörterbuchekurrenten artikeladkurrenten Zugriffspfaden oder zu den artikel- und wörterbuchekurrenten registeradkurrenten Zugriffspfaden. Da die Wörterbuchbenutzung, in deren Rahmen nichtmediostrukturelle wörterbuchekurrente äußere Zugriffspfade etabliert werden, wohl relativ selten auftritt, wird hier auf die Präsentation eines Typologieausschnittes verzichtet.

3.1.2 Zusammengesetzte äußere Zugriffspfade

Gegeben sei folgendes Beispiel:

(B 3-6: Zusammengesetzter äußerer Zugriffspfad/registeradkurrenter äußerer Zugriffspfad mit binnenkurrentem artikelorientiertem Anschlusspfad)

Werner benötigt bedeutungsverwandte Wörter zu dem Verb *herausfischen*; sein Benutzerziel ist mithin das Auffinden solcher Wörter. Er greift zum Dornseiff (2004) und führt mit dem Leitelement $\{herausfischen\}$ eine externe Registerzugriffshandlung erfolgreich aus, die ihn zu dem Registerzugang „**herausfischen**“ des einfachen akzessiven Registerintrags: „**herausfischen** 20.8 Aneignen“ führt. Damit hat Werner einen registeradkurrenten äußeren Zugriffspfad etabliert; um sein Benutzerziel zu erreichen, muss er sodann anhand der einfachen Registerangabe „20.8 Aneignen“, die wie eine reduzierte Verweisangabe funktioniert, das mediostrukturelle Leitelement $\{20.8 Aneignen\}$ erschließen und eine externe Verweisbefolgungshandlung ausführen, die ihn zu dem Sachgruppenartikel „**20.8 Aneignen**“ führt. Nach dem Zugriff auf diesen Artikeltitel als Eingang zum sachgruppenbezogenen Wörterbuchartikel hat Werner einen binnenkurrenten artikelorientierten Anschlusspfad etabliert, dessen Pfadende der Artikeltitel ist, und damit insgesamt einen zusammengesetzten äußeren Zugriffspfad, der zum Typ des registeradkurrenten äußeren Zugriffspfadestypes mit binnenkurrentem artikelorientiertem Anschlusspfad gehört. Das Kriterium dafür, dass im Beispiel B 3-6 nicht zwei unabhängige Zugriffspfade vorliegen, sondern nur ein zusammengesetzter äußerer Zugriffspfad, besteht darin, dass ein Benutzerziel alle pfadetablierenden Benutzungshandlungen determiniert.

Der Typ des zusammengesetzten äußeren Zugriffspfad es lässt sich wie folgt definieren:

(D 3-14: zusammengesetzter äußerer Zugriffspfad)

Ein *zusammengesetzter äußerer Zugriffspfad* ist ein äußerer Zugriffspfad, der aus einem adkurrenten äußeren Zugriffspfad als Anfangspfad besteht, an dem sich ein mediostruktureller binnenkurrenter äußerer Zugriffspfad anschließt, wobei die Bedingung gilt, dass alle pfadetablierten Benutzungshandlungen von nur einem Benutzerziel determiniert sind.

Das allgemeine Zugriffspfadbild für den Typ des registeradkurrenten äußeren Zugriffspfad es mit binnenkurrentem artikelorientiertem Anschlusspfad findet sich in Abb. 3-12.

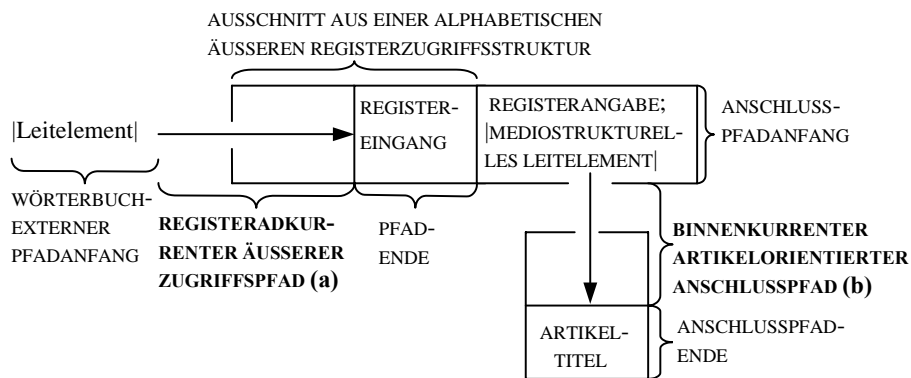


Abb. 3-12: Allgemeines Zugriffspfadbild für den Typ des registeradkurrenten äußeren Zugriffspfad es mit binnenkurrentem artikelorientiertem Anschlusspfad

Die meisten zusammengesetzten äußeren Zugriffspfade, die bei der Wörterbuchbenutzung etabliert werden, gehören zum Typ des registeradkurrenten äußeren Zugriffspfad es mit binnenkurrentem artikelorientiertem Anschlusspfad.

Bei der Benutzung von Spezialwörterbüchern sowie allgemeinsprachlichen Wörterbüchern mit reichhaltiger Umtextausstattung werden jedoch auch Zugriffspfade etabliert, die zu den anderen Typen gehören. Daher wird in Abb. 3-13 ein kleiner Ausschnitt aus einer Typologie von zusammengesetzten Zugriffspfad en präsentiert.

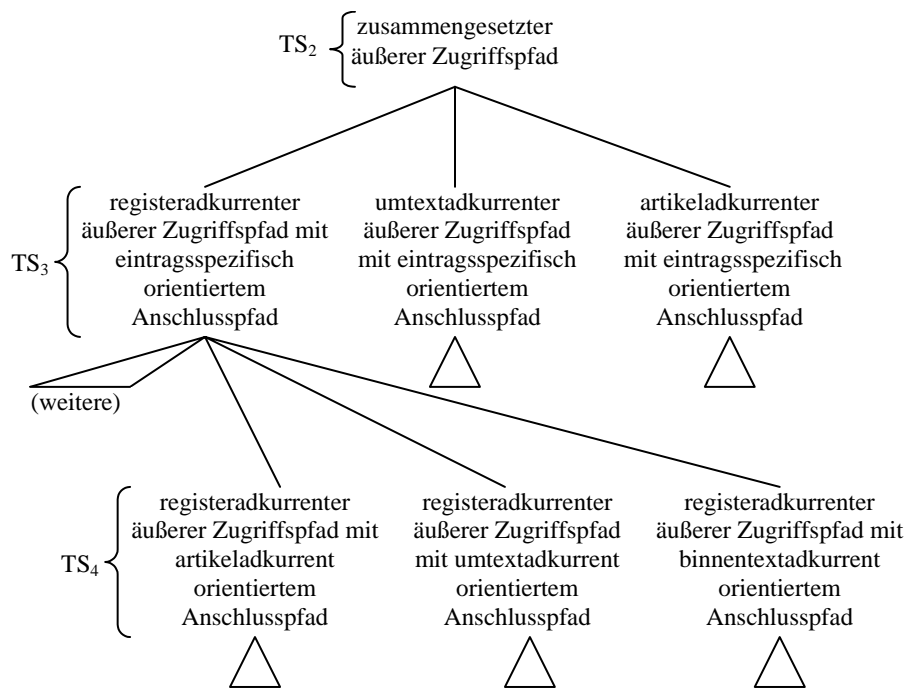


Abb. 3-13: Siebenter nichtkommentierter Typologigraph zu einem Ausschnitt aus einer Typologie von Zugriffspfaden

Zum Schluss dieses Abschnittes sei wenigstens erwähnt, dass für die Lern- und Konsultationswörterbücher (*sensu* Schierholz und Wiegand 2005 und Wiegand 2006a), zu den z.B. die „Wörterbücher zur Sprach- und Kommunikationswissenschaft“ [WSK] gehören, wenn diese zur wörterbuchinternen Fachwissensrecherche (*sensu* Wiegand 2006a) benutzt werden, vom Fachlexikographen empfohlene Zugriffspfadfolgen angegeben werden können, die von zusammengesetzten äußeren Zugriffspfaden unterschieden werden müssen, da sie nicht von *einem* Benutzerziel determiniert werden, sondern durch verweismotivierte Konsultationshandlungen (*sensu* Wiegand 1998: 408ff) verknüpft sind. Zugriffspfadfolgen lassen sich z.B. wie folgt angeben:

Zugriffspfadfolge I:

- (1) Umtextexkurrenter_{SE} synopseartikeladkurrent orientierter äußerer Zugriffspfad
- (2) Synopseartikelexkurrenter einzelartikeladkurrent orientierter äußerer Zugriffspfad

Zugriffspfadfolge II:

- (1) Einzelartikeladkurrenter äußerer Zugriffspfad

- (2) Einzelartikelexkurrenter synopseartikeladkurrent orientierter äußerer Zugriffspfad
- (3) Synopseartikelexkurrenter umtextadkurrent_{SE} orientierter äußerer Zugriffspfad

Der untere Index „SE“ besagt, dass es sich bei dem Umtext um die Systematische Einführung zu einem WSK-Band handelt, der nach Paragraphen gegliedert ist, an deren Ende jeweils eine feste Verweisposition situiert ist, in der auf die Synopseartikel verwiesen wird.

3.2 Akzessive Angaben und äußere Zugriffspfade

Im Folgenden soll gezeigt werden, dass und wie alle Typen von akzessiven Angaben, die zum Typologieausschnitt in Abb. 2-1 gehören, mit der Terminologie für äußere Zugriffspfade in einheitlicher Weise definiert werden können. Allerdings können (aus Platzgründen) nur einige Definitionen angegeben werden. Begonnen sei mit der Definition des Typs der extern akzessiven Angabe.

(D 3-15: extern akzessive Angabe)

Eine *extern akzessive Angabe* ist eine akzessive Angabe, die entweder über adkurrente äußere Zugriffspfade oder über registeradkurrente äußere Zugriffspfade mit binnenkurrentem eintragungsspezifisch orientiertem Anschlusspfad erreichbar ist.

Die drei in Abb. 2-1 angegebenen Untertypen des Typs der extern akzessiven Angabe lassen sich entsprechend wie folgt definieren:

(D 3-16: extern direkt akzessive Angabe)

Eine *extern direkt akzessive Angabe* ist eine akzessive Angabe, die nur über adkurrente äußere Zugriffspfade erreichbar ist.

(D 3-17: extern indirekt akzessive Angabe)

Eine *extern indirekt akzessive Angabe* ist eine akzessive Angabe, die nur über registeradkurrente äußere Zugriffspfade mit binnenkurrentem eintragungsspezifisch orientiertem Anschlusspfad erreichbar ist.

(D 3-18: extern direkt und indirekt akzessive Angabe)

Eine *extern direkt und indirekt akzessive Angabe* ist eine Angabe, die sowohl über adkurrente äußere Zugriffspfade als auch über registeradkurrente äußere Zugriffspfade mit binnenkurrentem eintragungsspezifisch orientiertem Anschlusspfad erreichbar ist.

Es ist klar, dass man bei den Definitionen von akzessiven Angaben nicht notwendigerweise auf die Terminologie für die Zugriffspfadtypen zurückgreifen

muss; vielmehr lassen sich die akzessiven Angaben auch unter Rückgriff auf die Typen von äußeren Zugriffsstrukturen sowie einige Typen von Benutzungshandlungen definieren. Solche Definitionen sind aber komplexer.

3.3 Verzeichnisexkurrente äußere Zugriffspfade

Es gibt Wörterbücher mit lexikographischem Inhaltsverzeichnis — wie z.B. Duden-⁶DUW —, und es gibt Wörterbücher ohne lexikographisches Inhaltsverzeichnis — wie z.B. Duden-⁴GFWB. Insbesondere bei den Ersteren fungiert die Paginierungsreihe als numerische äußere Zugriffsstruktur; sie heißt *Seitenzahlzugriffsstruktur*. Ihr direkter Zugriffsbereich ist die Menge aller paginierten Wörterbuchseiten. Zu jeder Seitenzahl als einem numerischen äußeren Zugriffstextelement gehört als Zugriffssektor eine Wörterbuchseite. Die Seitenzahlzugriffsstruktur ist eine mediostrukturelle äußere Zugriffsstruktur, so dass auf deren Zugriffstextelemente nur mit Hilfe mediostruktureller Leitelemente zugegriffen werden kann. Der erste Teil des lexikographischen Inhaltsverzeichnisses aus Steinhauer (2005) hat folgende Form:

Zur Bildung, Gliederung und Schreibung von Abkürzungen und Kurzwörtern	9
Im Register verwendete Kürzel	14
Verzeichnis der Abkürzungen und Kurzwörter	15
Abkürzungslisten:	
Deutsche Bundesländer (offizielle Zwei-Buchstaben-Kürzel)	82
Bundesministerien	82
Bundeswehr-Dienstgrade	83
Chemische Elemente	93
Diplomgrade	115
Doktorgrade	122
IATA-Codes deutscher Flughäfen	163

Ganz allgemein gilt, dass lexikographische Inhaltsverzeichnisse — wie auch die Register — als mediostrukturelle Komponenten eines Wörterbuchs zu gelten haben. Ein lexikographisches Inhaltsverzeichnis lässt sich am besten als eine geordnete Liste von reduzierten alphanumerischen Doppeladressenangaben verstehen. Die Reduzierung besteht darin, dass keine Verweisbeziehungsangabe existiert. Jede Inhaltsverzeichniszeile, wie z.B. „Bundeswehr-Dienstgrade 83“ besteht aus einer sprachlichen Unteradressenangabe (hier: „Bundeswehr-Dienstgrade“) und einer numerischen Hauptadressenangabe (hier: die Seitenzahl „83“). Mit der Unteradressenangabe werden meistens Umtexttitel oder Titel von eingelagerten Binnentexten genannt. Die Ordnung

der Liste wird durch die genannten Hauptadressen, also durch die Seitenzahlen, bestimmt: Geordnet wird nach steigenden Seitenzahlen (9, 14, 15, ... usw.). Ein lexikographisches Inhaltsverzeichnis liefert also dem Benutzer — wie ein Register — Fundortdaten, die als mediostrukturelle Leitelemente fungieren können. Im Unterschied zu einem Register weist ein lexikographisches Inhaltsverzeichnis aber keine äußere Zugriffsstruktur auf: Auf ein Inhaltsverzeichnis greift man nicht — unter Zuhilfenahme von vorgegebenen Ordnungsgesetzlichkeiten — zu, sondern man liest in ihm und kann gegebenenfalls auf eine Seite zugreifen, weil man eine intuitive Kenntnis der Ordnungsgesetzlichkeiten einer Zahlenreihe hat. Ein Benutzer, der dies tut, erbringt allerdings die gleiche oder zumindest eine sehr ähnliche kognitive Leistung wie ein Benutzer, der anhand eines alphanumerischen verweisvermittelnden Textsegmentes einen Verweis erschließt und diesem folgt. Es trägt daher zu einem vertieften systematischen Verständnis bei, wenn man ein lexikographisches Inhaltsverzeichnis als eine geordnete Menge von verweisvermittelnden Textsegmenten versteht. Man kann dann wie folgt argumentieren: Ein Benutzer-in-actu, der eine Verzeichniszeile liest, wie z.B. die Zeile „Diplomgrade 115“, und dann zur Seite 115 kommt, erschließt zuerst anhand der Verzeichniszeile, die als verweisvermittelndes Textsegment fungiert, einen Verweis, dem er daraufhin durch die Ausführung einer externen Verweisbefolgungshandlung folgt; dies geschieht dadurch, dass er eine externe Zugriffshandlung ausführt, die erfolgreich endet mit dem Zugriff auf die Seitenzahl „115“, die ein Element der Seitenzahlzugriffsstruktur ist; es folgt eine interne Zugriffshandlung, die bei „Diplomgrade“ erfolgreich endet. Ein Benutzer, der auf diese Weise z.B. vom Inhaltsverzeichnis zu einem Binnentexttitel gelangt, hat damit einen verzeichnisexkurrenten binnentextadkurrent orientierten äußeren Zugriffspfad etabliert; hinsichtlich der Wörterbuchseite als dem Zugriffssektor der Seitenzahl ist der etablierte Zugriffspfad inkurrent, nämlich wörterbuchseiteninkurrent, denn er wurde mit Hilfe einer Doppeladresse etabliert. Ein Benutzer, der vom Inhaltsverzeichnis zu einem Umtexttitel gelangt, hat einen verzeichnisexkurrenten umtextadkurrent orientierten äußeren Zugriffspfad etabliert, der ebenfalls wörterbuchseiteninkurrent ist. Die verzeichnisexkurrenten äußeren Zugriffspfade gehören zu den mediostrukturellen binnenkurrenten äußeren Zugriffspfaden. Für die, welche lexikographische Inhaltsverzeichnisse als Umtexte eines besonderen Typs begreifen, sind verzeichnisexkurrente äußere Zugriffspfade zugleich umtextexkurrente äußere Zugriffspfade.

3.4 Zugriffsumwege

In manchen Wörterbüchern sind die Schnittstellen von Benutzungshandlungen und Wörterbuchform nicht optimal gestaltet, sondern so, dass der Benutzer Zugriffsumwege in Kauf nehmen muss. Es gibt verschiedene Typen von Zugriffsumwegen, worauf ich hier nicht eingehen kann. Im Folgenden wird

nur ein Beispiel behandelt. Wir können hierzu auf den Binnentextabschnitt aus ²Herders Sprachbuch im Beispiel B 3-2 zurückgreifen. An dessen Schluss finden sich die beiden nichtusuellen binnentextorientierten Verweisangaben „□ Wortschatz“ und „□ Wortbildung“. In den Benutzungshinweisen mit dem Titel „Zum Aufbau des Sprachbuchs“ (S. VII–VIII) findet sich folgendes Textstück:

„Das Inhaltsverzeichnis auf S. XXVIII führt die Rahmenartikel in systematischer Reihenfolge auf.

Das Quadratzeichen (□) verweist auf den größeren sprachwissenschaftlichen Zusammenhang, in den der einzelne Begriff hineingehört; innerhalb der Rahmenartikel bezeichnet es den Titel des Artikels, in dem der Sachverhalt behandelt wird.“

Bei den Rahmenartikeln handelt es sich um 36 eingelagerte Binnentexte ohne äußere Binnentextzugriffsstruktur. Die beiden oben genannten nichtusuellen Verweisangaben stehen im Binnentext mit dem Binnentexttitel „Fach- und Sondersprachen“. Mit den beiden Verweisadressenangaben „Wortbildung“ und „Wortschatz“ werden Binnentexttitel genannt. Ein Benutzer, der auf diese zugreifen möchte, hat keine direkte Zugriffsmöglichkeit. Vielmehr muss er erst im Verzeichnis der Rahmenartikel im Wörterbuchvorspann nachschlagen und dort die Seitenzahl für die Seite suchen, auf der der Binnentexttitel steht und kommt erst dann auf diese Seite; der Benutzer muss also zuerst einen binnentextexkurrenten verzeichnisadkurrent orientierten äußeren Zugriffspfad etablieren und daraufhin als Anschlusspfad einen verzeichnisexkurrenten binnentextadkurrent orientierten äußeren Zugriffspfad. Eine Schnittstellengestaltung, die dem Benutzer einen Zugriffsumweg über das Verzeichnis in Form eines zusammengesetzten äußeren Zugriffspfades zumutet, ist aber unnötig und damit unprofessionell. Denn die Verweisangaben könnten einfach lauten „↗ Wortbildung, S. 328“ und „↗ Wortschatz, S. 697“. Oder aber man benutzt die alphabetische makrostrukturelle äußere Zugriffsstruktur auch als äußere Zugriffsstruktur für Binnentexte und ordnet die Binnentexttitel alphabetisch ein.

3.5 Innere Zugriffspfade

Während äußere Zugriffspfade immer über äußere Zugriffsstrukturen führen, führen innere Zugriffspfade immer über innere Zugriffsstrukturen. Außer den alphabetischen inneren Zugriffsstrukturen, die sich z.B. als positionsspezifische Zugriffsstrukturen in der Literaturposition von Fachwörterbuchartikeln finden, unterscheiden sich alle inneren Zugriffsstrukturen grundsätzlich von äußeren Zugriffsstrukturen, denn sie basieren auf dem Wissen von Benutzern über Artikelstrukturen (vgl. Wiegand 2005: 225ff). Benutzer können interne Zugriffshandlungen nur ausführen, wenn sie kognitiv über ihre individuelle interne Zugriffsstruktur verfügen. Der Pfadanfang von nichtmediostrukturellen artikelinternen Zugriffspfaden ist meistens die jeweilige lemmatische Angabe und das Pfadende die gesuchte Angabe. Bei erfahrenden Benutzern führen die

inneren Zugriffspfade meistens über die inneren Schnellzugriffsstrukturen, weil sich die Benutzer deren Elemente gut merken können. Der Pfadanfang von mediostrukturellen inneren Zugriffspfaden ist ein verweisvermittelndes Textsegment, mit dem eine Verweisbinnenadresse genannt wird. Im Unterschied zu den eintragsübergreifenden Zugriffspfaden, die binnenkurrent verlaufen, ist der Pfadverlauf bei inneren Zugriffspfaden z.B. artikel- oder umtext-intern. Eine Typologie von inneren Zugriffsstrukturen ist relativ unergiebig.

3.6 Übersicht über die Typologie von Zugriffspfaden in Printwörterbüchern

In der Abb. 3-14 findet sich eine Übersicht über die Typologie von Zugriffspfaden:

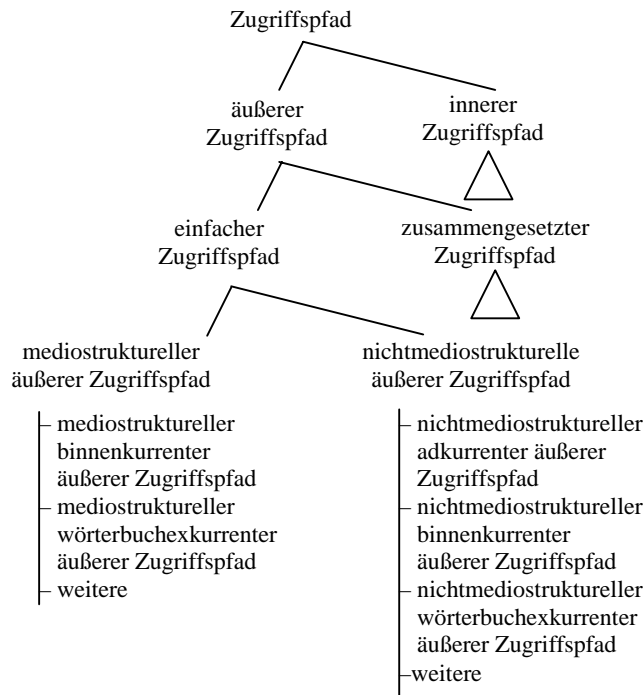


Abb. 3-14: Achter nicht kommentierter Typologigraph: Übersicht zur Typologie von Zugriffspfaden

3.7 Schlussbemerkung

Die in diesem Beitrag erstmals in einer einheitlichen Perspektive zusammenhängend vorgestellte Terminologie für Zugriffspfade in Printwörterbüchern und die zugehörigen lexikographietheoretischen Begrifflichkeiten leisten bei

der Evaluation der Datenakzessivität von Printwörterbüchern und bei der Darstellung des mediostrukturellen Profils eines Wörterbuchs sowie bei dessen systematischer Planung gute Dienste. Dies wird in einem späteren Beitrag dargestellt, der an den vorliegenden anschließt und in dem eine formale Sprache für die Darstellung von Zugriffspfaden und damit für die Darstellung von sogenannten Verweissystemen präsentiert wird.

Danksagung

Werner Wolski danke ich sehr für eine hilfreiche Durchsicht des Manuskripts und Christina-Doreen Reinhard für dessen Herstellung.

Literatur

- DCE = 1984:** *Dictionary of Contemporary English. Ein umfassendes einsprachiges Wörterbuch für Schule und Hochschule.* London: Longman Group Ltd.
- Dornseiff, Franz.** 1934. *Der deutsche Wortschatz nach Sachgruppen.* Berlin/Leipzig: Walter de Gruyter.
- 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-⁶DUW =** Dudenredaktion (Hrsg.). 2007. *Deutsches Universalwörterbuch.* 6., überarb. und erw. 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-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 A.G.
- ²Herders Sprachbuch =** 1971. *Herders Sprachbuch. Ein neuer Weg zu gutem Deutsch.* Freiburg [etc.]: Verlag Herder KG.
- 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.
- Steinhauer, Anja.** 2005. *Duden. Das Wörterbuch der Abkürzungen. Rund 50 000 nationale und internationale Abkürzungen und Kurzwörter mit ihren Bedeutungen.* 5., vollst. überarb. und erw. Aufl. Mannheim [etc.]: Bibliographisches Institut & F.A. Brockhaus AG.
- 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.** 2002. Altes und Neues zur Mediostruktur in Printwörterbüchern. *Lexicographica* 18: 168-252.
- Wiegand, Herbert Ernst.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, Herbert Ernst.** 2005. Über die Datenakzessivität in Printwörterbüchern. Einblicke in neuere Entwicklungen einer Theorie der Wörterbuchform. *Lexikos* 15: 196-230.
- Wiegand, Herbert Ernst.** 2006. Das Lern- und Konsultationswörterbuch. Ein neuer Fachwörterbuchtyp am Beispiel der *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK)*. *Lexikos* 16: 205-221.
- Wiegand, Herbert Ernst.** 2006a. Die „Wörterbücher zur Sprach- und Kommunikationswissenschaft“ (WSK) und ihre Benutzungsmöglichkeiten im Fach Deutsch als Fremdsprache. Dimova, Ana, Vida Jesenšek und Pavel Petkov (Hrsg.). 2006. *Zweisprachige Lexikographie und Deutsch als Fremdsprache. Drittes Internationales Kolloquium zur Lexikographie und Wörterbuchforschung. Konstantin Preslavski-Universität Schumen, 23.–24. Oktober 2004*: 1-35. Germanistische Linguistik 184-185. Hildesheim [etc.]: Georg Olms.
- Wiegand, Herbert Ernst.** 2007. Mittel zur Darstellung textueller Wörterbuchstrukturen von Printwörterbüchern. Ein Beitrag zur Methodologie der systematischen Wörterbuchforschung. *Lexicographica* 23. [Erscheint.]
- WNED** = 1993. *Webster's New Encyclopedic Dictionary*. New York: Black Dog & Leventhal Publishers Inc.

Creating a Single South African Keyboard Layout to Promote Language*

Dwayne Bailey, *Translate.org.za* (dwayne@translate.org.za)

Abstract: In this case study, a description is given of a keyboard layout designed to address the input needs of South African languages, specifically Venda, a language which would otherwise be impossible to type on a computer. In creating this keyboard, the designer, Translate.org.za, uses a practical intervention that transforms technology from a means harming a language into one ensuring the creation and preservation of good language resources for minority languages. The study first looks at the implications and consequences of this missing keyboard, and then follows the process from conception, strategy, research and design to the final user response. Not only are problems such as researching the orthographies, key placement and keyboard input options examined, but strategic objectives such as ensuring its wide adoption and creating a multilingual keyboard for all South African languages are also discussed. The result is a keyboard that furthers multilingualism and ensures the capturing of good data for future research. Finally it is a tool helping to boost and bolster the vitality of a language.

Keywords: KEYBOARD, MULTILINGUALISM, VENDA, AFRIKAANS, TSWANA, NORTH-ERN SOTHO, ZULU, SOURCE, FREE SOFTWARE, LAYOUT

Opsomming: Die skep van 'n enkelvoudige Suid-Afrikaanse toetsborduitleg om taal te bevorder. In hierdie gevallestudie word 'n beskrywing gegee van die ontwerp van 'n sleutelborduitleg vir die hantering van die insetbehoefes van Suid-Afrikaanse tale, veral Venda, 'n taal wat andersins onmoontlik op 'n rekenaar getik sou kon word. Deur die skep van hierdie sleutelbord gebruik die ontwerper, Translate.org.za, 'n praktiese ingryp wat tegnologie verander van 'n middel wat 'n taal benadeel tot een wat die skep en bewaring van nuttige taal-hulpbronne vir minderheidstale verseker. Die studie kyk eers na die implikasies en gevolge van hierdie ontbrekende sleutelbord, en volg dan die proses van konsepsie, strategie, navorsing en ontwerp tot die uiteindelijke gebruikersreaksie. Nie alleen word probleme soos die navorsing van die ortografieë, sleutelplasing en sleutelbordinsetteuses ondersoek nie, maar ook strategiese doelwitte soos die versekering van 'n wye aanvaarding daarvan en die skep van 'n meertalige sleutelbord vir alle Suid-Afrikaanse tale word bespreek. Die resultaat is 'n sleutelbord wat meertaligheid bevorder en die vaslegging van nuttige gegewens vir toekomstige navorsing verseker. Uiteindelik is dit 'n werktuig wat help om die lewenskragtigheid van 'n taal te versterk en te steun.

Sleutelwoorde: TOETSBOARD, VEELTALIGHEID, VENDA, AFRIKAANS, TSWANA, NOORD-SOTHO, ZOELOE, BRON, GRATIS PROGRAMMATUUR, UITLEG

* This article was presented as a paper at the Eleventh International Conference of the African Association for Lexicography, organised by the Tshivenda National Lexicography Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5-7 July 2006.

1. Introduction

Translate.org.za is involved in the localisation of free and open source software (FOSS)¹ into the eleven official languages of South Africa. The organisation delivered the first office productivity suite (word processor, spreadsheet, presentation program) in these languages in 2005, followed by the Mozilla Firefox web browser and the Mozilla Thunderbird email program in 2006.

During this process Translate.org.za soon discovered that Venda translators could not physically type the five extra characters² needed in Venda orthography on their computers. Most translators chose simply to ignore these characters or to follow convoluted processes to insert them. Even more surprising was that language bodies were seriously considering revision of the language's orthography to adapt to the limitations of technology.

Of great concern to Translate.org.za were some of the methods used to circumvent this problem. These included printing and manually adding diacritics, adding characters that looked like the correct ones, and using methods that printed correctly and looked correct but were not actually the correct characters. For Translate.org.za the use of the correct characters are essential. As a character cannot be pencilled onto a piece of software, another character or even one that looks correct cannot be used. It was soon realised that Venda translators around the country were creating parallel texts that were of no use for research, that these documents looking good on paper could never be published in electronic format, and that in fact their actions would have a long-term detrimental effect on Venda as, for instance, Translate.org.za is unable to produce spell checkers using Government texts with the diacritics missing.

Translate.org.za's own needs and the belief that the problem could be addressed lead to the research on and development of the South African keyboard layout.

Most users take a keyboard for granted; it simply works and is very reliable. If it breaks, you buy a relatively inexpensive replacement. What most people think of as the keyboard is the physical keyboard derived from the computer's typewriter heritage. A keyboard has a layout, a definition of the positioning of the keys. In South Africa, the dominant layout is the US English layout. But you will most probably have encountered a British English layout and would also have seen layouts for French, German and other languages.

Although you will have a computer keyboard with a defined physical layout, you also have a software layer that defines this layout for the computer. This usually maps one-to-one to the physical keyboard. However, as it is software driven, this can be changed to redefine the behaviour of the physical keyboard. Thus it can be defined that when you press S you get š. In this process, a new layout is being created that could in time be made into a physical keyboard. What Translate.org.za has developed is a South African keyboard layout, creating a new software mapping to the physical keyboard.

Although, from a software side, the creation of the keyboard layout was

relatively simple, the strategic decisions and research into the design took much longer to formulate and test. The team devoted attention to the following aspects:

- (a) Language investigation — What additional characters did each language need?
- (b) Decision on one or multiple language layouts — Should each language have a keyboard layout or could/should one be created to cover all South African languages?
- (c) Determination of the layout — How should the keys be positioned for ease of use and adoption, considering the different types of users (touch typists and two-finger typists)?
- (d) Consideration of the keyboard — Should a physical keyboard be created?

The team settled on combining all languages into one keyboard layout with three different layouts catering for different types of users. Instead of a physical keyboard, an overlay was developed that allows any existing keyboard to be converted into a South African keyboard.

From the design phase, it was soon realised that Venda still has the problem of displaying the characters if the computer does not have a Venda-capable font installed. To overcome this, Translate.org.za extended the existing free DejaVu font to include the missing Venda characters and combined this with the keyboard software.

The final step has been to validate and test the layouts. Fortunately a number of translators have tested and evaluated the keyboards and given constructive feedback. Three layouts are not ideal as only one design would have been preferred. But for now, Translate.org.za is testing and refining with the hope that only one will remain. In the long term, this could evolve into a standard South African keyboard layout.

Translate.org.za, over and above its software localisation effort, continues to undertake work which touches on the broader issues of computers and South African languages. These include amongst others: designing spell checkers, creating and validating collation sequences and creating locale information. Translate.org.za has received funding from various organisations, including the Department of Communications, the International Development Resource Centre (IDRC), The Open Society Institute (OSI) and the Shuttleworth Foundation.

2. What lead to the creation of this keyboard?

2.1 The history of Translate.org.za

Translate.org.za is a South African non-profit organisation focused on the localisation, or translation, of free and open source software (FOSS) into South

African languages and on the creation of FOSS localisation tools such as translation memory tools, translation editors and translation management software.

Under its first mandate, Translate.org.za released the first word processor in a South African language in 2004 when the Zulu, Northern Sotho and Afrikaans versions of OpenOffice.org were introduced. Subsequently a version in all eleven official languages has been created and was released in 2005. In addition to this, the Mozilla Firefox web browser and the Thunderbird email program have been completed in all languages. Translate.org.za has added spell checkers and now a keyboard to make this software fully multilingual.

Towards its second mandate, Translate.org.za is a partner in the WordForge project which is developing FOSS localisation tools.

Translate.org.za believes that concrete action is the only way to foster language pride, challenge negative language perceptions and build a multilingual society. The keyboard is a good example of such action.

2.2 Language pride

When championing language pride what could be more debilitating than technology that cannot correctly work in a language? The perception is then that the language is in some way inferior or backward.

2.3 Observing usage patterns

Since Translate.org.za translates software into all eleven official languages, it is thus also translating into Venda. During training sessions with translators, patterns and problems were noticed that had not been seen in other languages. Venda translators were struggling to enter text in their language. The problem had been ignored for so long that most translators were surprised when questions were asked about their ability to type characters for Venda diacritics.

None of the solutions used by translators addressed the root of the problem and none of them were transportable between applications. One solution involved: type, print, then pencil in diacritics. It was clear Translate.org.za could not accept pencilled-in characters. Many other issues were found with the workarounds. For instance, solutions would work in Microsoft Word, but not in Microsoft Excel. Or the solutions would work well, yet not in Windows 98.

3. The present situation

3.1 Workarounds used by translators

The following patterns were found which place translators into two groups, namely those who let technology master them and those who create some workaround even if it is convoluted and inferior.

3.1.1 Those who are mastered by technology

The way they cope with these computer problems also reflects on the translators' respect for their own language. They will do one of the following:

- (a) Ignore the diacritics — simply type using the 26 letter alphabet.
- (b) Add the diacritics later — type, print or pencil in the missing characters.
- (c) Choose characters that look suitable but are not the correct ones — N dot below (ñ), instead of N dot above (ñ).

The consequence of any of these actions is data unfit for further use or research. Here is a short list of possibilities lost with these actions:

- (a) You cannot reuse any of the translations in a translation memory.
- (b) Researchers cannot use any of this data for any language analysis.
- (c) You cannot repurpose a document, e.g. create a printed copy and a copy to publish on the Internet.
- (d) You cannot collaborate, i.e. you cannot email a document for review.

Thus every day translations are being created that could be enhancing the status of languages. Yet every day work is being produced that is unfit for advancing the languages in any way.

3.1.2 Those who have mastered technology

Fortunately the second group is larger than the first, people who do care about correctly representing their language even if they have invented very convoluted ways of doing so. Unfortunately they are also very averse to change. This can be a problem with some of the solutions they have designed as these do not create valid output. The following are the methods they use:

- (a) Menu driven: Select, Insert → Special Character, select the Venda diacritic from a dialogue box.
- (b) Special key combinations: Associate special characters to a special key combination in their word processor. These options have been removed in the newer version of Microsoft Office.
- (c) Autocorrect: Create special autocorrect sequences, e.g. typing ^n will create ñ.
- (d) Macros: Create special Office macros to add the characters. (It is uncertain whether these were valid Unicode characters or whether they merely looked correct.)

3.2 Revisiting orthographies

This, also known as the workarounds developed by linguists, has been referred to as mechanical imperialism,³ changing the language to meet the shortcomings of technology. The extreme of this would of course be to change the complete writing style.

Orthographies do need to be revisited and updated from time to time. But this should be based on adaptations of the language, not on the shortcomings of the tools used to transcribe the language.

4. The role of a keyboard in the promotion of multilingualism

Translate.org.za is very focused on practical interventions. It is felt that not enough of these interventions are effected in the effort to promote multilingualism.

The keyboard interventions were:

- (a) Simple — These are relatively easy to implement.
- (b) Practical — These are easy to demonstrate.
- (c) Helpful — These simplify the operation for users.
- (d) Unifying — These bring languages together.

5. The keyboard in the computer process

Although this can seem quite technical, it is often important that language users obtain insight into how a computer understands their input. This helps them to identify certain types of problems and comprehend their causes, and their solutions.

5.1 Character encoding (Unicode)

Each alphabetic character is represented by a number, also known as a code point. A number of code points creates an encoding. A group of encoded characters is known as a character set or a character set encoding. In the early days of computing, there was only limited space for character storage. Thus there were different encodings for different language groups: Western European, Eastern European, Japanese, etc. The problem with this approach is that if the wrong encoding is specified, the text will be garbled, i.e. an A in the Latin alphabet will appear as another character in Cyrillic, because they have the same encoded number.

However, this was solved with the creation of Unicode, which has enough space to encode every character of every living language in the world. In fact,

Venda can only be stored using the Unicode character set. An added advantage of Unicode is that since every character has a unique number, it is possible to mix languages in a single document. So it is now possible to write an English paper describing an Arabic translation of a Chinese text.

Many older computers cannot handle Unicode correctly⁴. This would explain why you cannot type Venda on a Windows 98 machine, as it is not fully Unicode compliant. Certain applications, by implementing their own system to manage Unicode, are able to give the illusion of compliance but only in the application, not across the system. This, however, would still not address the ability to type on a keyboard.

5.2 Fonts

With the ability to store the characters, the ability to see them is now also needed. For this there are fonts and each font contains a number of glyphs, a glyph containing the drawing instructions for one character. The glyph is associated with a code point. Thus if Venda text is stored on your computer, the computer will know the code point of that character. By also knowing the encoding, the computer determines that that code point represents N with a dot above (ñ) in the Unicode encoding. Since the encoding of the font is also known, the correct glyph can be pulled from the font and thus the correct character can be drawn and displayed. With the absence of the correct glyph, you will see a small square box (□) or a question mark (?). This is an indication that you do not have the correct font.

5.3 Keyboard layout

In more complex languages, such as Thai and Khmer, there are issues related to inputting of the language into the computer that are much more advanced than the simple keyboard used by Latin-based languages, thus South African languages are unaffected.

Once a character can be stored and displayed, the keyboard allows the same character to be entered. The keyboard layout ensures that when a key or a set of keys are typed, the computer can determine what code point the user requested. With the code point, the correct character can be displayed and stored. As a user, you simply see that the key you press produces the correct character in your word processor.

5.3.1 Dead keys

The name "dead key" is a hold-over from the days of typewriters. If you typed a dead key on a typewriter, the carriage would not move. Thus the second character would be typed in the same space. So, for instance, typing ^ followed by E would create ê. Modern computer keyboards can be made to emulate this behaviour.

5.3.2 AltGr

In many languages where there are not enough keys, the right Alt key, also called AltGr, allows a keyboard to have many more characters, e.g. AltGr + E could create ê.

6. Strategic options

There were a number of options that could have been pursued in the design of the keyboard and certain constraints that caused a move in a certain direction.

6.1 Goal

This is a brief summary of the pursued goals. These evolved over time as it was realised and learnt what was possible and what was not workable.

- (a) Unobtrusive — A keyboard must be created that could exist without anyone knowing it was there. If it does not get in your way, you are unlikely to want to remove it, yet it remains in place in case you or someone else requires it.
- (b) Simple — The key sequences must be easy to remember with logical associations.
- (c) One-off — One keyboard must be designed for all languages, so that any language can be typed on the same keyboard. If users need to exchange keyboard layouts, they would most likely not adopt a new one, but revert to their old US English layout.
- (d) Cheap — The keyboard must be inexpensive so that nobody has an excuse not to use it.

6.2 Design

6.2.1 Keystroke analysis

At first, it was felt that keystroke analysis would be a good approach as it would create an optimal keyboard. By monitoring which keys are used most frequently, the high frequency characters could be placed on keys which are easy to reach. This would create the best keyboard for speed of typing in a given language. The current QWERTY keyboard layout was designed to slow people down when using the first mechanical typewriters; it was not designed for speed.

In consultation with linguists working on Nigerian languages, it was soon realised that people are hard to change. Once you can use a QWERTY key-

board, you are unlikely to use another. Such a demanding change would fail on a number of the pursued goals.

6.2.2 Keyboard format

How to present the keyboard to users?

6.2.2.1 A physical keyboard

The option of a physical keyboard would allow the localisation of keys such as Enter, Shift, Caps Lock, etc. However, it is costly with a relatively small market. The question though is this: Do you read or need to read your keyboard? Once users know the key associations, they no longer need the hints from a physical keyboard.

This option fails dramatically as an inexpensive goal. It would also make it difficult to adopt. Would someone supply you with a Venda keyboard when it would add a considerable sum to the price of your computer? In the first stages, it was decided that this was not the best strategy, although it is still being investigated as an option.

6.2.2.2 Labels

Another option is simple stick-on labels that can work with an existing English US keyboard. The labels are not essential but allow relatively easy conversion of any existing keyboard into a South African keyboard. Production costs, however, are still high.

6.2.2.3 Configuring the keyboard

The keyboard can be presented as a software-only solution. For now, this is what Translate.org.za is doing, supplying the software layout that allows users to configure and operate the keyboard without any physical crutches such as a keyboard or keyboard labels.

6.2.3 Layout

The decisions on layout revolved around what keyboard style to use:

- (a) Remap unused keys.
- (b) Use AltGr.
- (c) Use dead keys.

With remapping of keys, a key or character that is unused in a language is used for a different character. For example, if the character C is unused in the lan-

guage, it can be remapped to another character such as š. As most South African computer users need to type English, this was, however, not an option.

With each of the last two possibilities, AltGr and dead keys, there is the decision around which keys to associate with each diacritic character. In the use of dead keys, the decision concerns which key should be used to represent the caron or the dot above. In the use of AltGr, it is to be decided which letter to associate with the character that contains the diacritic. For instance: N needs to be used twice in Venda for N with a dot above (ñ) and N with a circumflex below (ñ). It must be decided which character should occupy the N, i.e. AltGr + N, and where the other character should be placed.

In all these decisions, the layout needs to remain logical, easy to use and easy to remember.

7. Implementing the South African keyboard

7.1 What characters are needed

This would seem to be one of the simplest tasks. However, it proved to be quite difficult, because it was necessary to ensure that everything had been covered. The problems encountered were changes in orthography, and remnants of old orthography.

It was discovered that, although orthographies had changed, the old orthography remained in, for instance, people's names. Thus it had not completely disappeared, and there would be instances in which the old orthography was needed such as population registration and legal contracts. Over time, this will become less of a problem.

In researching the characters, a number of people were consulted, but the vagueness of most answers made it difficult finding a reliable resource. There seemed to be confusion about what was old and what was new, and what was required for the current orthographies. Dictionaries, grammars and various study materials were consulted. Some presented problems as they included many tone marks not used in the written form of the language. Thus extracting a simple list of characters was troublesome.

In addition to this, old orthographies in texts such as the Tswana Bible were still prevalent. In each instance of a new authoritative source that presented new characters, it had to be investigated whether this was current or old orthography.

7.2 Understanding the targets and tools

The primary operating systems targeted are Linux and Windows. The first prototype was produced on Linux, but subsequently most of the development was done on Windows as the tools are more easily used for prototyping keyboard layouts. The layouts will be ported back to Linux once they are stable.

Windows proved to be problematic in its own way. Since Windows did not start supporting Unicode properly until Windows 2000, it took much research to discover that proper keyboard input for Venda on the Windows 9x platform will never be attained.

On Windows, the Microsoft Keyboard Layout Creator (MSKLC) is employed. This is a useful tool as it is quite quick to prototype layouts. It was frustrating that the tools do not actually compile a usable install package which Translate.org.za could handle itself.

On Linux, the process of developing the layout is quite cumbersome with little documentation of a clearly powerful system. Work is under way to improve the keyboard handling on Linux's windowing system. Keyboard layouts will have to be proved before undertaking the porting task.

No work has been done on creating a layout for the Apple Mac range of computers. The process is relatively simple but the need is not as great as the Windows keyboard layout.

7.3 The final outcome

As it was found that there are different sets of users, it was actually not possible to make one design, which would have been preferred. Thus these keyboards with the following user profiles were created.

AltGr	Infrequent user, computer used by multiple people, public terminal, user is not a touch typist
Intl	User types at some speed and finds the AltGr combinations awkward
Super	Touch typist

It is felt, however, that the AltGr layout will prevail as it is simple to understand and use, even though it is not ideal for speed.

7.3.1 One keyboard

It was hoped that a single keyboard could be created. Afrikaans with its many diacritics was a concern as these quickly fill up the layout. But since there is very little overlap between Venda, Northern Sotho and Afrikaans, it was found that all characters could be catered for on one keyboard layout.

7.3.2 Key associations and how these were arrived at

Logical associations are needed between keys and their diacritics. On the AltGr keyboard, the circumflex and the caron are the dominant symbols. Although

the diacritics are not the same characters, they do look the same. Venda has a circumflex below, Afrikaans a circumflex above and Northern Sotho and Tswana S caron. Thus this was chosen as the dominant character. So if you type AltGr + N, you will get N with a circumflex below ($\underset{\sim}{n}$). AltGr + S results in S caron (š).

In Afrikaans, the diaeresis was chosen as the dominant character as it appears on all vowels, while the circumflex is needed on only some of the vowels.

Other diacritics occurring on the same key were placed as close to the key as possible and, if possible, on a key of related shape. Thus AltGr + M will give N with a dot above ($\overset{\cdot}{n}$). M was chosen because it is shaped like N.

In Afrikaans, the circumflexes proved a problem as the keyboard around U, I and O became rather crowded. But since the AltGr keyboard is designed to be used with labels, it was concluded that this pattern can be learnt.

The Intl (International) keyboard was a compromise in that many expert Afrikaans computer users make use of it. It is also an easy layout to learn as there are logical associations for diacritic characters. All diacritics are made with dead keys, so this is not a non-intrusive keyboard. A person unfamiliar with the layout will be frustrated. The US Intl keyboard layout was merely taken and South African characters added. Thus $\hat{\ } + S$ will give S caron (š), $\hat{\ } + N$ will give N circumflex below ($\underset{\sim}{n}$) and $\hat{\ } + E$ will give E circumflex (\hat{e}).

The last keyboard was designed to meet the needs of touch typists. When broken down, there is very little need to have multiple dead keys. In this layout, a relatively comfortable position was chosen for the dead keys. Comfort rather than associations was considered. When you look at the characters that are needed, no language makes use of more than two diacritic marks (ignoring of course acute and breve in Afrikaans). Since it was already found in the AltGr that there is a common visual association around caron and circumflex, it was decided to treat these similarly. Thus in the super keyboard, if you type $< + S$, you get S caron (š), $< + E = E$ circumflex above (\hat{e}), while $> + N = N$ dot above ($\overset{\cdot}{n}$) and $> + E = E$ diaeresis (\ddot{e}).

The keys ; and ' handle acute and breve as dead keys because they are easily located. It is still uncertain whether these two are the best choice as dead keys are wanted that are not too invasive and these effect the single quotation mark.

7.3.3 The need for fonts

Venda needs fonts and not everyone has them (unless of course you think typing documents in Tahoma is readable). Most people working with Microsoft Office will have support in Arial Unicode if they have installed them. There are a number of fonts that will cover the Venda characters; most notable are those from SIL (that has just begun to release its fonts under terms that would make them freely distributable).

In order to make sure that everyone has access to at least one set of fonts

that are freely distributable, Venda characters were added to the DejaVu font. This gives a serif, sans serif and monospace font. Strangely enough, users are satisfied with such a limited choice. Choice is, however, very much a part of language pride.

8. Outcomes

Extensive testing has now been performed with a number of translators. The following, sometimes unexpected, results were obtained.

8.1 Simple can be hard

The AltGr layout meets all of the intended goals, but according to the feedback from translators it was at times awkward to use. AltGr + Shift + N can be so awkward that some users in fact just reverted to the MS Word shortcut associations or to the Insert menu.

8.2 The difficulty of change

For many people changing from the familiar to something new was impossible to even consider. Thus they abandoned the keyboard quite easily.

8.3 No ownership

This is a phenomenon that open source developers have always found rather baffling. If there is a problem, an open source user tries to figure it out and get it fixed. Translators, who are probably a good reflection of average Windows users, would simply consider a problem as a fault of the software and thus abandon it as inferior, ignore the problem or work around it.

9. Further developments

Looking forward, there is still work to be done on the keyboard. The layouts should remain relatively unchanged, but there are a number of additional items that are considered:

- (a) Porting to Linux and Mac — AltGr works on Linux but the other layouts are still to be migrated so that they can be used on these platforms.
- (b) Collaboration with SABS and CSIR on standardisation — Discussions have been started that may lead to some form of standardisation work on the layouts.
- (c) Creation of a real keyboard — A physical keyboard needs to be created for those who believe firmly in multilingualism.

- (d) Investigation of a Venda-only layout — Due to the awkwardness of the layout, further investigation should be done on the concept of a Venda-only layout that could work without dead keys.
- (e) Completion of orthographies — For researchers in the language field, the International layout should be expanded to cover all characters that have been used in all orthographies, past and present, of all South African languages.
- (f) Installation of more fonts — Venda-speaking people should be given the ability to use any font with their characters.

10. Conclusion

Translate.org.za has succeeded in its goals of creating a keyboard layout for South Africa. It is one practical step in raising the issue of language pride in South Africa and in restoring dignity to languages ignored by computers.

Further testing, checking, adoption and adaptation need to be performed and, most importantly, stability achieved before this has a chance of becoming a South African standard. But in the long run, it is hoped that people will be using keyboards that work in their mother tongue.

Acknowledgements

Translate.org.za would like to thank the Afrikaans users who tested the various incarnations of the keyboard and who criticised the AltGr keyboard for its awkwardness. Although more feedback is needed, the Venda translators who have been working with this keyboard must be congratulated for starting what will hopefully become a concerted fight against mechanical imperialism.

Notes

1. FOSS is an area software development that focuses on the rights of users of software. It has the right to distribute, modify, use and share modifications. It is a rapidly growing area of software development that is well known to most through the Linux operating system. OpenOffice.org, Firefox and Thunderbird are all part of FOSS.
2. These characters are: ḑ, Ḑ, ṛ, ṛ̣ and ṛ̤.
3. Many thanks to Ronald Madiba from whom I first heard this phrase.
4. There is a Unicode layer for Windows 9x. However, this is only a translation layer on top of the existing codepage system of the 9x series. Thus, if your characters do not appear in any codepage, they will not work in the Unicode layer. Since Venda's characters do not appear in any codepage, they cannot be managed by a Windows 9x computer.

Dictionary Writing System (DWS) + Corpus Query Package (CQP): The Case of *TshwaneLex*

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),*
and

Guy De Pauw, *CNTS — Language Technology Group, University of Antwerp, Belgium; and School of Computing and Informatics, University of Nairobi, Kenya (guy.depauw@ua.ac.be)*

Abstract: In this article the integrated corpus query functionality of the dictionary compilation software *TshwaneLex* is analysed. Attention is given to the handling of both raw corpus data and annotated corpus data. With regard to the latter it is shown how, with a minimum of human effort, machine learning techniques can be employed to obtain part-of-speech tagged corpora that can be used for lexicographic purposes. All points are illustrated with data drawn from English and Northern Sotho. The tools and techniques themselves, however, are language-independent, and as such the encouraging outcomes of this study are far-reaching.

Keywords: LEXICOGRAPHY, DICTIONARY, SOFTWARE, DICTIONARY WRITING SYSTEM (DWS), CORPUS QUERY PACKAGE (CQP), TSHWANELEX, CORPUS, CORPUS ANNOTATION, PART-OF-SPEECH TAGGER (POS-TAGGER), MACHINE LEARNING, NORTHERN SOTHO (SESOTHO SA LEBOA)

Samenvatting: Woordenboekaanmaaksysteem + corpusanalysepakket: een studie van *TshwaneLex*. In dit artikel wordt het geïntegreerde corpusanalysepakket van het woordenboekaanmaaksysteem *TshwaneLex* geanalyseerd. Aandacht gaat zowel naar het verwerken van onbewerkte corpusdata als naar geannoteerde corpusdata. Wat het laatste betreft wordt aangetoond hoe, met een minimum aan intellectuele arbeid, automatische leertechnieken met succes kunnen worden ingezet om corpora voor lexicografische doeleinden aan te maken waarin de woordklassen expliciet worden vermeld. Alle stappen van de redenering worden geïllustreerd met gegevens uit het Engels en Noord-Sotho. De instrumenten en technieken zelf zijn echter allemaal taalafhankelijk, waardoor de veelbelovende resultaten van deze studie verrijkend zijn.

Sleutelwoorden: LEXICOGRAFIE, WOORDENBOEK, SOFTWARE, WOORDENBOEK-AANMAAKSYSTEEM, CORPUSANALYSEPAKKET, TSHWANELEX, CORPUS, CORPUSANNOTATIE, WOORDKLASSETAGGER, AUTOMATISCHE LEERTECHNIEKEN, NOORD-SOTHO

1. DWS & CQP: What are these?

Metalexigraphers, dictionary makers and software engineers are constantly seeking better and faster ways to produce today's dictionaries. While software tools to assist the compilation of reference works are becoming ever more advanced, complete digital solutions are also increasingly put directly into the hands of the lexicographers themselves. Most professional dictionary houses that compile dictionaries in-house use (at least) two sets of tools: a dictionary writing system (DWS) on the one hand, and some sort of corpus query package (CQP) on the other. A team of local IT gurus will then typically ensure the transition of data between these two systems. It is very rare that the two systems are truly integrated, and if they are, a publisher-specific setup was designed. Until recently, no off-the-shelf packages — that anyone could acquire, for the compilation of any type of dictionary, for any (number of) language(s) — combining both a DWS and a CQP, were available. Since the release of the TshwaneLex Suite 3.0 in June 2007, however, this is now a reality.

Being a South African product, TshwaneLex is well-known in South Africa, as it is the system of choice for the eleven PanSALB-sponsored National Lexicography Units (NLUs), is in use at all the major local dictionary publishing houses (including OUP Southern Africa, Pharos and Macmillan SA), and copies of TshwaneLex are also found at virtually all South African universities. Its sister application TshwaneTerm, designed for the management of terminology, is popular with various government departments, including the South African Police Service (SAPS). Worldwide, there are currently over three hundred users of the TshwaneLex Suite (which now bundles TshwaneLex and TshwaneTerm).

In this article, the main issues that are relevant when integrating a CQP with a DWS are analysed. Specific attention goes to the current status quo at the NLUs. The strong and weak points uncovered are then used as a point of departure for the discussion of the built-in corpus functionality of TshwaneLex. A major leap forward is subsequently suggested in the form of the utilisation of integrated part-of-speech tagged corpora, and a method to obtain these for under-resourced languages is worked out for Northern Sotho as an illustration. In conclusion, a fully-working system is presented.

2. DWS >< CQP: The status quo

Before discussing the setup at the South African NLUs, it is useful to briefly look at two of the most advanced setups in the world, namely that of the 'IMS Textcorpora and Lexicon Group' at the University of Stuttgart (Heid et al. 2007), and that of the users of the 'Sketch Engine' (Kilgarriff et al. 2007). Over the course of many years, the Stuttgart team has developed the following linguistic resources and tools:

- Lexicons
- Tools for automatic text analysis and corpus annotation
- Retrieval and extraction tools
- Linguistically annotated text corpora
- Linguistic engineering standards

Whenever they have to compile new dictionaries and/or analyse, revise or update existing dictionaries, they employ these linguistic resources and tools, but none of these is integrated into their DWS, and new, ad hoc bits of code have to be written every time to link the two.

Arguably the most powerful CQP currently available is the Sketch Engine. In the words of Adam Kilgarriff: "The Sketch Engine (SkE, also known as Word Sketch Engine) is a Corpus Query System incorporating word sketches, grammatical relations, and a distributional thesaurus. A word sketch is a one-page, automatic, corpus-derived summary of a word's grammatical and collocational behaviour." The lexicographers thus have these one-page summaries in front of them, but still need to manually copy-and-paste data from the SkE into their DWS.

Less advanced, but still within the same framework, the South African lexicographers currently use WordSmith Tools (WST, Scott 2007) as their CQP.

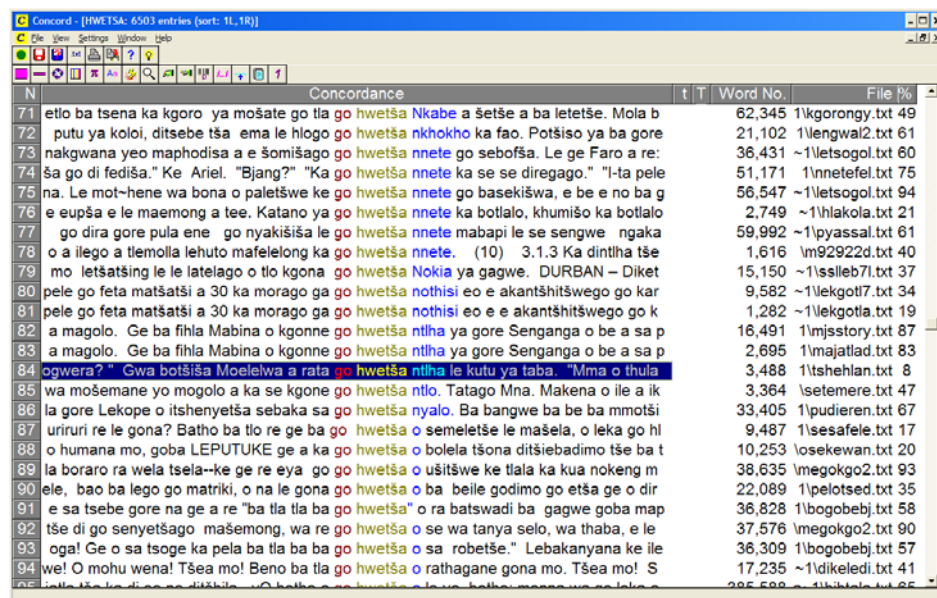


Figure 1: The corpus query package (CQP) WordSmith Tools, as used at the Northern Sotho National Lexicography Unit (NLU)

Figure 1 shows a screenshot of the use of a corpus at the Northern Sotho NLU, where corpus lines for the verb *hwetša* 'find; get; obtain' are shown, sorted one to the left as primary sort, and one to the right as secondary sort. For each corpus line, the source can be seen in the rightmost column.

In this case, the lexicographers have selected corpus line 84 to illustrate one of the senses of *hwetša*, and a straightforward copy of the highlighted line in WST allows them to paste this data into TshwaneLex, following a toggle between programs. Corpus examples are only adapted minimally, and after clean-up and adaptation, the result is as seen in Figure 2, which is a screenshot of a search for *hwetša* in the current online version of this NLU's dictionary (Mojela et al. 2007).

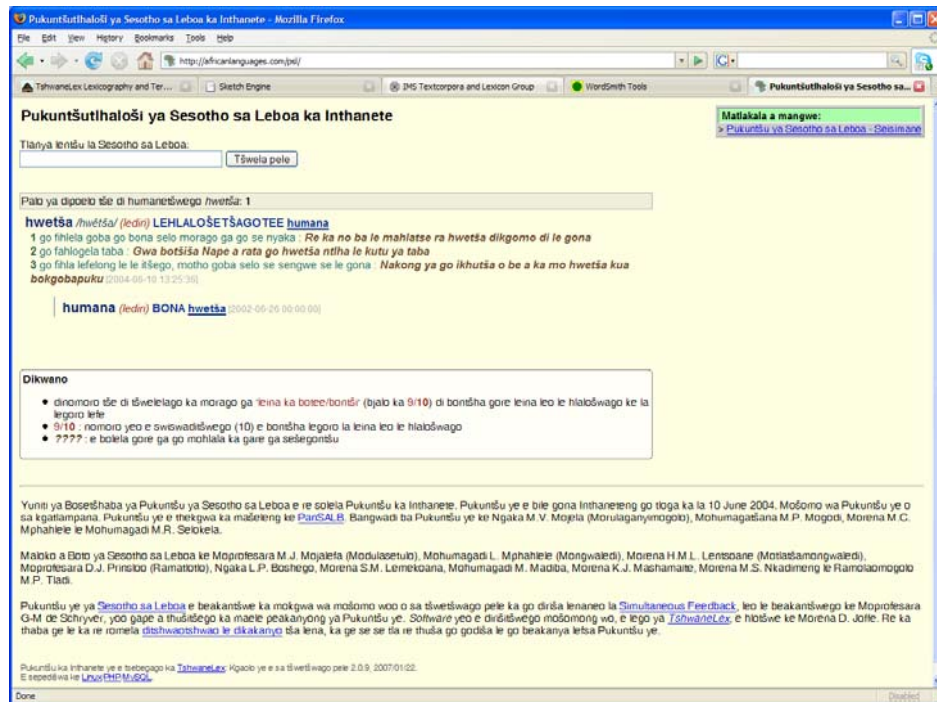


Figure 2: A search for *hwetša* in the online monolingual Northern Sotho dictionary [Note that the example for sense 2 corresponds to line 84 in Figure 1]

This setup whereby two separate pieces of software are used, on the one hand TshwaneLex and on the other WST, works well, but the fact that data has to be manually selected and copied over between programs is cumbersome, and there are also some encoding issues to be circumvented (for Northern Sotho, for instance, typing in the 'š' in the CQP is often a problem on standard South African keyboards).

3. DWS + CQP: Wish list

What one would want to have is a single computational environment as dictionary compilation system, but without the loss of all the main advantages of using a CQP like WST. If one studies the functions that are needed on a daily basis for the compilation of a dictionary, then one arrives at the following minimum package:

- The ability to search for any word, combination of words, or multiple words simultaneously;
- The above in combination with wildcards;
- A function to easily sort the lines, according to the search node, as well as the word immediately to the left and immediately to the right;
- An indication of the occurrence (frequency) of the searched-for item;
- A function that allows for easy sampling; absolutely crucial for frequent words (such as function words);
- An indication of the source a particular corpus line comes from;
- An easy way to transfer data from the corpus analysis environment to the dictionary compilation environment.

All of the above are available in WST, although many clicks are needed for some of these basic functions that are needed in lexicography. (But then, WST is not primarily meant to assist lexicographers.)

Two further requirements may be added to the wish list. Firstly, a tool to encrypt the corpus data, so as to protect it from theft, is often desirable. Indeed, for all resource-scarce languages, a considerable amount of effort went into the creation of the now-available corpora, and while one wants all the dictionary compilers to be able to access the corpus data, one simultaneously does not want the data to be 'just out' there, 'in the open' so to say. Secondly, and implicit in the description of the South African setup so far, it would be much more meaningful to be able to work with corpora that are annotated for parts of speech, rather than having to work with the current 'raw' (and thus linguistically 'blind') corpora.

4. DWS + CQP: TshwaneLex's F6 — Part 1: Working with raw corpus data

Figure 3 shows an overall picture of the TshwaneLex interface, here with data for a bidirectional Northern Sotho–English dictionary (in Linked View mode). For introductory, detailed as well as technical discussions of TshwaneLex, the reader is referred to the URL mentioned in the References of this article (Joffe et al. 2007). The focus here is on the built-in corpus functionality only, which is conveniently accessed by pressing the function key F6. In the interface, the corpus tool is therefore also the sixth tab of the attributes window (to be found at the bottom-right of each dictionary side). In the screenshot of Figure 3, the

Northern Sotho to English side of the dictionary can be seen on the left-hand side, while the English to Northern Sotho side may be seen on the right-hand side. Two different test corpora have been loaded, texts 1 to 4 for Northern Sotho, and four works by well-known authors for English.

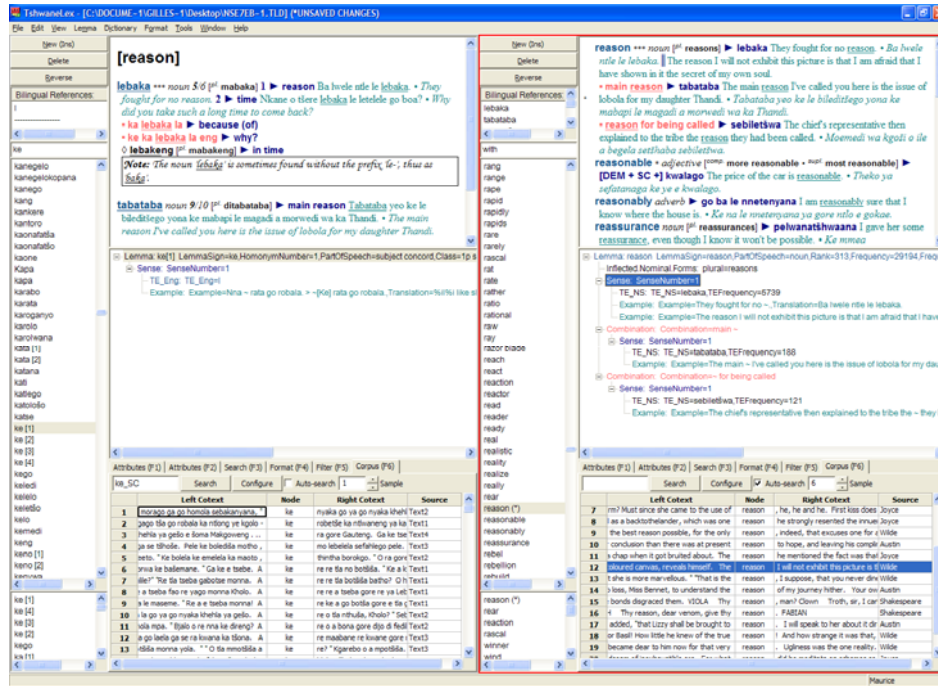


Figure 3: An overall picture of the TshwaneLex interface, with a focus on F6 for both sides (i.e. the built-in corpus tool)

All aspects of the 'minimum requirements' listed in the wish list of Section 3 have been implemented, and this seamlessly. Figure 4 zooms into the bottom-right corner of Figure 3, in this case the English corpus. In the search box one can type in any string of characters and press "Search" to initiate a search through the loaded corpus files. Regular expressions can also be used in this input box, which means that wild cards and more are easily available. To limit the number of results, sampling was performed in Figure 4, here with a random selection of each sixth line in the corpus. Actually, nothing was typed into the search box in Figure 4, as the "Auto-search" option had been ticked. This means that the corpus is searched automatically for the lemma sign of any article that one is working on. To further streamline the dictionary compilation process, one can simply stand on one (or more) corpus line(s) and press Ctrl+F7: the highlighted sentence(s) will automatically be copied over to the sense one is working on. In Figure 4 this was done for line 12, the result of which can

be seen in Figure 5. Compilation can then proceed as usual, which in the case of a bilingual dictionary means, amongst others, that one will translate the chosen example(s).

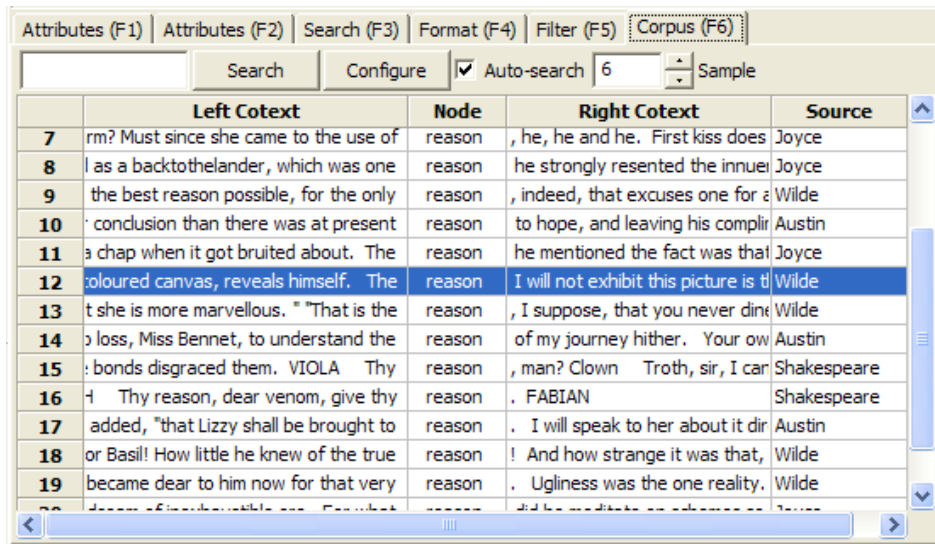


Figure 4: Using F6 (TshwaneLex's corpus tool) during dictionary compilation — Part 1: Selecting corpus lines for inclusion at the sense one is working on

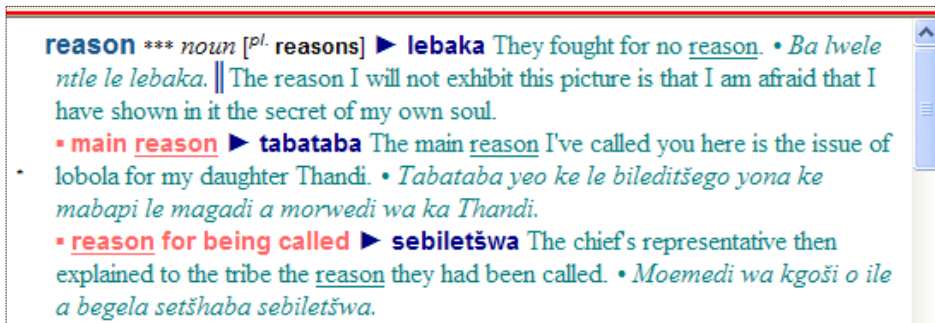


Figure 5: Using F6 (TshwaneLex's corpus tool) during dictionary compilation — Part 2: Automatic transfer of entire sentences to the sense one is working on

Pressing the "Configure" button under F6 launches the dialog shown in Figure 6. From it, it can be deduced that functionality has indeed been included that allows for corpus data to be encrypted. TshwaneLex remembers corpus settings, so that corpora are automatically loaded whenever one starts work. With this dialog, the files that make up the corpora can also easily be changed.

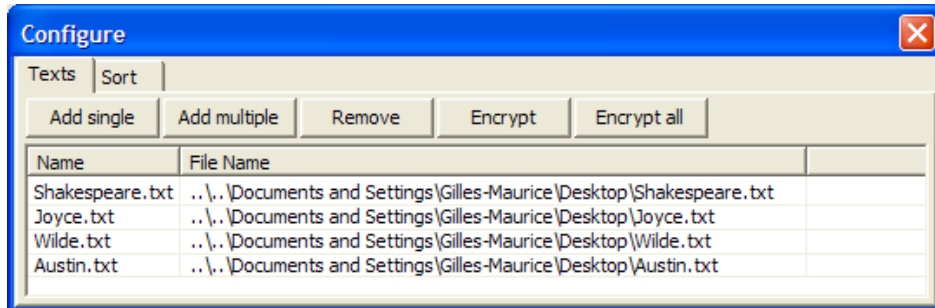


Figure 6: The F6 (TshwaneLex's corpus tool) dialog

What is still outstanding from the wish list, then, is the possibility to handle corpora that are marked up for parts of speech. There are two aspects that need attention here. Firstly there is the issue to annotate a corpus, and secondly there is the issue of how to handle such annotated corpora in TshwaneLex. Given corpus annotation has never been pursued in full in South Africa (nor for any African language for that matter, bare Swahili, cf. Hurskainen 1999), we feel that the lexicographic community would gain much from a presentation that shows how this can be accomplished with a minimum of effort. The next section illustrates this.

5. Intermezzo: Part-of-speech tagging under-resourced languages

Research in the field of Natural Language Processing (NLP) is inconceivable without the availability of digital resources developed by the lexicographic community. All too often though, this natural and apparent link between the two related communities is ignored and true collaborative research efforts are few and far between. In the experiments described in this section, we show how one can directly exploit lexical resources in a cost-effective development of language technology applications. As a case study, we describe how dictionary data can be used to create an accurate part-of-speech tagger for Northern Sotho with a minimal amount of manual effort. Once such a tagger is available, *better* dictionaries can be made *faster*. One thus comes full circle.

5.1 Data-driven part-of-speech tagging

In recent years, part-of-speech (POS) taggers have become established language technology tools for most of the major Indo-European languages, such as English, French, Dutch and German. POS-taggers are an essential component for many commercial NLP applications, such as machine translation or text mining, serving as one of the most valuable disambiguation steps in the processing

chain. Assigning the parts of speech to the words in a sentence indeed provides invaluable morphological and syntactic information that percolates to all levels of linguistic analysis.

While historically POS-taggers have most often been constructed by hand-crafting rule-based systems, the last two decades have seen a definite shift to more robust, data-driven approaches. Rather than capturing the solution to the POS-tagging problem in a set of manually designed rules, these data-driven methods exploit large, manually annotated corpora, to train statistical methods and machine learning approaches that can automatically induce the solution to the disambiguation problem.

The advantages of the data-driven approach are many. First and foremost, it is inherently language independent: the construction of a POS-tagger for a given language or domain is relatively straightforward, provided there is annotated data available for it. This contrasts the construction of rule-based POS-taggers, which is typically a labour-intensive process, requiring the involvement of linguistic experts and producing taggers that are not portable to other languages. Furthermore, data-driven taggers have been shown to outperform hand-crafted taggers on most types of language data, not only in terms of development costs, but also in terms of coverage and robustness (Kupiec 1992).

While the advantages of data-driven tagging are obvious, they are restricted by the fact that they require annotated data for a given language. While annotated resources are abundantly available for languages such as English and French, this is not the case for the majority of the world's languages. Consequently, this impedes the automatic development of language technology tools for the under-resourced languages.

Many researchers assume that data-driven approaches to NLP require the availability of hundreds of thousands of annotated tokens. A relatively new area of NLP research is now investigating how more modest-sized data sets can also bootstrap language technology for resource-scarce languages. In this section, we describe how we can build a relatively accurate data-driven POS-tagger for the resource-scarce language of Northern Sotho on the basis of a manually annotated corpus of only 10 000 words.

5.2 From lexical database to annotated corpus

Typically, the construction of a POS-tagged corpus requires a carefully designed tagging protocol, which involves the definition of a well-rounded tag set and meticulously stipulated tagging guidelines. While this is a valuable exercise in its own right, we hypothesise that most aspects of the protocol can be dynamically constructed and refined during actual annotation, provided there is a strong lexical backbone that can bootstrap the annotation process.

The TshwaneDJe HLT Northern Sotho lexical database provides exactly this backbone, since it includes for each lemma its frequency information and its possible POSs. From this database we can extract the top 5 000 words and

list the POS-tags associated with those words. The typically Zipfian distribution of language tokens in a corpus ensures the tag set extracted from the lexical database has sufficiently large coverage. Furthermore, the annotation environment easily allows minor adjustments to be made to this tag set, should it be unable to adequately cover some tokens. This on-the-fly approach enables the organic construction of a consistent tag set, grounded in linguistic, corpus-based evidence. The complete Northern Sotho tag set induced from the lexical database and further refined during annotation, is shown in Addendum 1.

The annotators were asked to tag a corpus of literary texts. The data was tokenised and imported into a spreadsheet, listing one word per row. The TshwaneDJe HLT Northern Sotho lexical database then further allowed us to guide and speed up the manual annotation efforts: for over 90% of the words in the corpus, the lexical database can provide the annotator with a 'shortlist' of possible tags for each word, thereby minimising manual input.

The procedure is illustrated in Figures 7 and 8, which are screenshots of the annotation environment in the spreadsheet. Column B in Figure 7 contains the word to be tagged, while columns D, E, F, etc. provide the possible tags for the word, as retrieved from the lexical database, and ordered according to overall frequency of occurrence. Words with more than one possible tag, thus ambiguous words, are highlighted in green, indicating the annotator needs to disambiguate this word. Words associated with just one tag are not highlighted, although the annotator can still change the tag, should it not be correct.

	A	B	C	D	E	F	G	H	I	J
6145	6145	<utt>								
6146	6146	Ga		LOCp	HRTp	NEG	PC			
6147	6147	ke		SC	COPp	AUX_V	AGTp			
6148	6148	ye		DEM	V					
6149	6149	go		SC_ind	SC	LOCp	OC	CP15		
6150	6150	sutha		SC						
6151	6151	tabeng	SC							
6152	6152	ya	SC_ind		V	PC				
6153	6153	ka	TMPp	IMPp	SC	LOCp	AUX_V	PRO_pos	POT	INSp
6154	6154	ya	unknown	V		PC				
6155	6155	gore	V+eng	DNJ						
6156	6156	ga	V+go	CP	HRTp	NEG	PC			
6157	6157	re	V+ng	SC	OC	V				
6158	6158	reke		V						
6159	6159	dijo		N						
6160	6160	.		Punc						
6161	6161	</utt>								

Figure 7: Spreadsheet containing the initial material for the annotator

Disambiguation is done as follows: if the tag in Column D is correct, nothing needs to be done and the annotator can proceed to the next word. If it is incorrect, the cell needs to be deleted and the annotator checks the tag in Column E. If this tag is correct, the annotator can move on to the next word. If it is not, the cell is deleted and Column F is considered and so on and so forth. If the correct tag is not provided by the lexical database in any of the columns, the annotator can use the drop-down box in Column C, which contains all the tags of the tag set (see Addendum 1). Post-processing will only retrieve the leftmost tag, which

means annotators need not spend time deleting incorrect tags to the right of the correct tag.

Unknown words are highlighted in grey. These are words that fall beyond the scope of the top 5 000 words in the lexical database. For these words, the annotator has to choose a tag from the drop-down box in Column C, or, if quicker, he/she can simply type the required tag in the empty cell in Column C. 'New' words that are likely to recur may be tagged in one go: once the annotator has provided the correct POS-tag for such a new word, he/she may sort Column B, copy-and-paste the tag for other occurrences of the same word, and then restore the original order, by sorting the data back using the indices provided in Column A. Once tagged, the sample from Figure 7 is as shown in Figure 8.

	A	B	C	D	E	F	G	H	I	J
6144	6145	<utt>								
6145	6146	Ga				NEG	PC			
6146	6147	ke		SC	COPp	AUX_V	AGTp			
6147	6148	ye			V					
6148	6149	go						CP15		
6149	6150	sutha	V							
6150	6151	tabeng		N+ng						
6151	6152	ya				PC				
6152	6153	ka						PRO_pos	POT	INSp
6153	6154	ya				PC				
6154	6155	gore		CONJ						
6155	6156	ga				NEG	PC			
6156	6157	re		SC	OC	V				
6157	6158	reke		V						
6158	6159	dijo		N						
6159	6160	.		Punc						
6160	6161	</utt>								

Figure 8: Spreadsheet containing the material annotated by the annotator

The contents of the drop-down box, that is the tag list, are defined on a separate worksheet. This tag list can be dynamically adjusted, should a new tag need to be created for a particular word. Spreadsheet functionality allows for the drop-down box in the annotation worksheets to automatically reflect the adjustments.

Despite the availability of dedicated annotation tools, using a spreadsheet for this annotation task has some significant advantages. Installation is trivial and most computer-literate users are familiar with this type of software, so that the learning curve for the annotators is favourable. While annotation is an unlikely use of a spreadsheet, the cell-based approach can significantly speed up an annotation task such as POS-tagging.

Practical limitations prevented us from annotating the entire dataset. Restricted to a total annotation time of a mere 10 person-hours, the design of the annotation environment just described nevertheless maximised the amount of annotated data. After post-processing the data, we had a manually tagged corpus of more than 10 000 words, in a format ready to be used as training material for a data-driven tagger. A sample of this is shown in Figure 9.

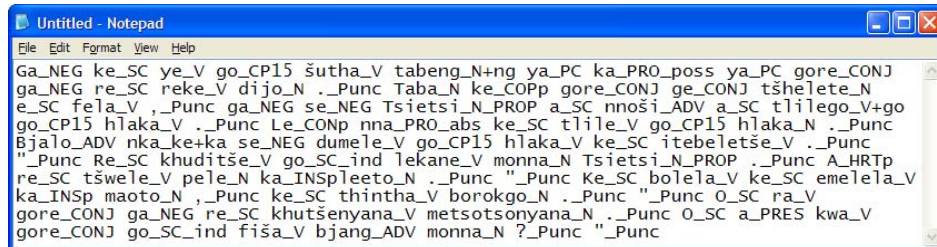


Figure 9: UTF-8 encoded text-only version of the POS-tagged corpus

While a 10 000-word-tagged corpus is indeed modest, compared to the million-word corpora available for English (Marcus et al. 1993), the experiments described in the next section will show that even a small annotated data set can yield an accurate data-driven POS tagger.

5.3 Maximum Entropy Tagging

The general idea behind data-driven POS-tagging is that annotated data, as displayed in Figure 9, inherently encode the solution to the disambiguation problem of POS-tagging. Rather than having a linguist explicitly specify that a word is to be tagged with some POS-tag in a particular context, the data-driven approach tries to automatically induce this type of disambiguation from annotated data.

In recent years, a large number of data-driven POS-tagging toolkits have become freely available, such as MBT (Daelemans et al. 2003), MXPOST (Ratnaparkhi 1996), TnT (Brants 2000) and SVMTool (Giménez and Márquez 2004). Given the availability of annotated data for some language, all of these tools become a viable option to construct a POS-tagger (De Pauw et al. 2006). While there are significant differences in the way these respective data-driven methods implement the solution to the problem, they all have in common that they try to 'mimic' the behaviour of the manual annotators, by trying to capture linguistic patterns using statistical and/or symbolic means.

The particular approach used in the experiments described in this article is based on the machine learning technique of Maximum Entropy Learning (Berger et al. 1996). This technique has previously shown to obtain state-of-the-art results for many languages, including Swahili. Rather than the stock maximum entropy tagger, MXPOST (Ratnaparkhi 1996), we used a self-constructed POS-tagger, called MaxTag, which acts as a front-end to the general machine learning package Maxent (Le 2004). It has the advantage of providing fast processing times and being more robust in handling morphologically rich languages.

MaxTag takes as its input POS-tagged data and extracts for each word in the corpus a number of features that are possibly relevant to the disambiguation problem. Just as in most other data-driven taggers, the features considered by MaxTag are a mixture of contextual and orthographic information. Ortho-

graphic features try to capture the morphological aspects of the word to be tagged, while contextual features describe the syntactic context of the word within the sentence.

Contextual features¹

- **FW**: the word itself
- **FT**: a single token representing all possible tags for the focus word
- **W-1**: the previous word
- **W+1**: the next word
- **T-1**: the preceding tag
- **T+1**: a single token representing all possible tags for the next word

Orthographic features

- **P1**: the first letter of the word
- **P2**: the first two letters of the word
- **P3**: the first three letters of the word
- **S1**: the last letter of the word
- **S2**: the last two letters of the word
- **S3**: the last three letters of the word
- **CAP**: the presence of a capital

For our tagging experiments, we recorded a limited context, considering only one word before and after the word to be tagged. While one might be tempted to record a substantially larger context size, there is a trade-off: the larger the contexts that are being considered, the less linguistic evidence one can find in the data, making it harder for the data-driven taggers to generalise over the data. Particularly on small data sets such as the one for Northern Sotho, the sparse data problem restricts us to using a more limited context. We illustrate the extraction of features for the following example sentence:

Ke_SC a_PRES eletša_V_Punc

For each word in the corpus an 'instance' is extracted that contains both contextual and orthographic features. Each instance is then associated with the POS-tag for that word. For the example sentence above, this gives us the following four instances shown in Table 1.

Once MaxTag extracts one instance for each word in the annotated data, the Maxent machine learner consequently processes all of these instances. It observes the association of features with a particular tag and tries to optimally estimate the 'predictive' power of a particular feature. It will for example try to infer from instance (1) that a focus word (FW) 'Ke' at the beginning of a sentence (cf. "W-1=#") is likely to be associated with the tag "SC".

Table 1: Four 'instances' for the sentence "*Ke_SC a_PRES eletša_V . _Punc*"

Instance	Tag
(1) ['W-1=#', 'T-1=#', 'FW=Ke', 'FT=SC_COPp', 'W+1=a', 'T+1=SC_PRES_PC_DEM_OC_HRTp', 'P1=K', 'S1=e', 'P2=Ke', 'S2=Ke', 'CAP']	SC
(2) ['W-1=Ke', 'T-1=SC', 'FW=a', 'FT=SC_PRES_PC_DEM_OC_HRTp', 'W+1=eletša', 'T+1=V', 'P1=a', 'S1=a']	PRES
(3) ['W-1=a', 'T-1=PRES', 'FW=eletša', 'FT=V', 'W+1=.', 'T+1=Punc', 'P1=e', 'S1=a', 'P2=el', 'S2=ša', 'P3=ele', 'S3=tša']	V
(4) ['W-1=eletša', 'T-1=V', 'FW=.', 'FT=Punc', 'W+1=#', 'T+1=#', 'P1=.', 'S1=.']	Punc

On the basis of these observations, the Maxent machine learner will try to construct a statistical model that optimally relates features to classes. It does this in an iterative movement, by attempting variations of the parameters in the statistical model and evaluating the tagging accuracy of the respective models on the annotated data. When the most optimal settings are eventually found, the resulting model will be able to tag new, previously unseen data, thereby roughly mimicking the tagging behaviour of the original annotators.

While tagging new data, both contextual and orthographic information is taken into account. In our evaluation, we will make a distinction between tagging accuracy on known words versus unknown words. The former are words that occur in the annotated data, and for which the tagger has observed some linguistic evidence. The latter are previously unseen words, that do not occur in the lexical database, nor in the annotated corpus. For these, orthographic features prove particularly useful, as they attempt to encode morphological information.

5.4 Experimental results

To evaluate the performance of the tagger, we performed a ten-fold cross validation experiment. This means we randomly distribute the sentences of the full data set over ten partitions and run ten experiments: in each experiment a different partition is used as the testing set, while the other nine partitions are used as the data to train the data-driven tagger. By comparing the tags output by the tagger to those in the original annotation, we can estimate the accuracy of the tagger on previously unseen data. Doing this ten times further reduces the risk of accidentally evaluating on an artificially (un)favourable training-test set partition, providing trustworthy experimental results.

We compare the results of the MaxTag method to that of a baseline. The baseline method is defined as the simplest solution to the problem. For known

words, the baseline method uses a unigram approach, always selecting the tag that is most frequently associated with a particular word in the training data. Unknown words are invariably tagged as nouns by the baseline method, since that is the tag most often associated with low-frequency words. The baseline method achieves a reasonable, but unspectacular tagging accuracy on unseen data, as can be seen from Table 2. Particularly the score for unknown words (about 8% of the test data) is underwhelming.

Table 2: Accuracy scores for the baseline method and MaxTag (all values in %)

	Known	Unknown	Total
Baseline	75.8	35.1	73.5
MaxTag	95.1	78.9	93.5

The MaxTag method achieves a significant increase on all accounts, particularly for unknown words. This amounts to a total tagging accuracy of 93.5%. Given the restricted size of the annotated corpus, this result is very encouraging, especially since the baseline indicates that the POS-tagging problem for Northern Sotho is far from trivial.²

5.5 Towards large POS-tagged corpora

The experiments described in this section have resulted in a POS-tagger that is able to tag unseen data with an accuracy of almost 94%. This opens up the possibility of semi-automatic annotation, where the data-driven tagger provides a first annotation, which is consequently checked by human annotators. This not only speeds up corpus annotation itself, but also ensures more consistent annotation throughout the corpus. This is an especially important point when working in a multi-annotator environment. Subsequent annotation efforts will undoubtedly greatly benefit from the Northern Sotho tagger developed in the context of this article.

While 93.5% tagging accuracy is an encouraging result, this is still not up to par with data-driven taggers for English (Van Halteren et al. 2001) or Swahili (De Pauw et al. 2006), achieving near-human type of tagging accuracies. This is undoubtedly due to the limited size of the annotated corpus.

We therefore ran some learning curve experiments to investigate the effect of the quantity of data on the quality of the resulting tagger. To this end, we isolated a single evaluation set and used the remaining data to incrementally train POS-taggers. We started out with a POS-tagger trained on just $\frac{1}{10}$ of the available training data (roughly 1 000 words) and added $\frac{1}{10}$ of the training data in each subsequent experiment. The result of this experiment can be found in the learning curve graph, displayed in Figure 10.³ The graph shows that the learning curve is still linear and that we can still gain quite a bit of tagging accuracy by collecting more annotated data. Future research will therefore con-

centrate on the semi-automatic development of new annotated data. The more than encouraging experimental results show that the Northern Sotho version of MaxTag can provide an invaluable tool in this endeavour.

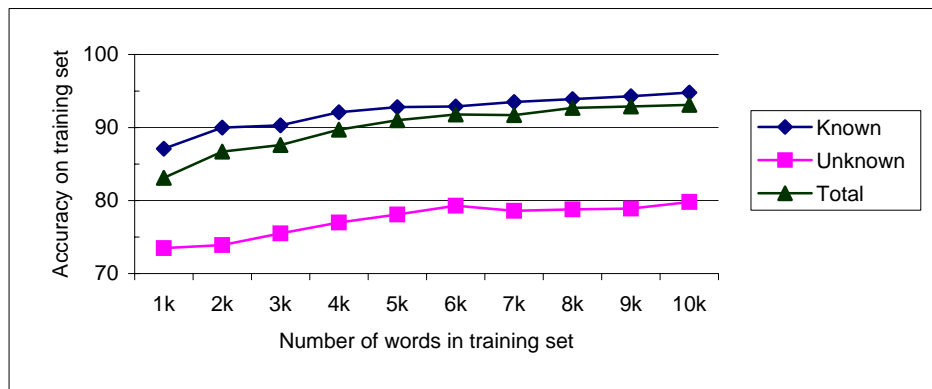


Figure 10: Graph for learning curve experiments

6. DWS + CQP: TshwaneLex's F6 — Part 2: Working with POS-tagged corpus data

From Section 5 we see that POS-tagged corpora are now within reach for both well-resourced and under-resourced languages. In this last section we briefly show how working with POS-tagged corpus data can significantly improve the use of corpora during dictionary compilation. The most obvious, and most straightforward use is that automated searches can now take the part of speech of the article one is working on into account. In Northern Sotho, there are for example four different homonymous *fela*, one being an adverb ('only; just'), another a conjunction ('but'), another a verb ('finish; come to an end'), and lastly an auxiliary verb ('usually; continuously'). With the "Auto-search" option enabled, as illustrated in Figure 11, a lexicographer working on the adverbial form will only be presented with those cases in the corpus tagged as adverbs, greatly simplifying the task to select an appropriate example. This is especially true for function words, where manual disambiguation between homonyms during dictionary compilation can be very taxing. In general, patterns are also much quicker to spot. So a search for "nyaka_AUX_V" will immediately reveal that this form is virtually always followed by a main verb in the infinitive, which may prompt the lexicographer to add a Usage Note in this regard.

On a next level, TshwaneLex's corpus tool also allows lexicographers to search for particular 'word-POS pairs', as can be seen in Figure 12 (which is an enlargement of the bottom-left corner of Figure 3).

Needless to say, all of these may be combined with regular expressions during a search, which means that one can now truly find what one is looking

for on the one hand, but also that one can now also more easily pinpoint linguistically-informed patterns in the data, which can then be described lexicographically.

The screenshot shows a software interface for working with POS-tagged corpora. On the left is a list of search results for the word 'fela', including various inflected forms like 'fega', 'fegelwa', 'fegelwana', 'fehla', 'fekese', 'fela [1]', 'fela [2]', 'fela [3]', 'fela [4]', 'fele', 'felegetša', 'felela', 'felele', 'felelela', 'feleletša', 'feleletše [1]', 'feleletše [2]', 'felelwa', 'feletša', 'feletše', 'feletšwe', 'felo', 'felwana', 'fentše', 'fentšwe', 'fenya', 'fenywa', 'fepa', 'feta', 'fete', 'fetela', 'fetetša', 'fetile', 'fetiša', 'fetišetša', 'fettleka', 'fetnna', 'fela [1]', 'fela [2]', and 'fe'. The main window displays detailed information for the lemma 'fela[1]'. It lists several senses with their part of speech (adverb, conjunction, verb) and provides example sentences in Setswana with English translations. A note states: 'Note: The auxiliary verb stem 'fela' is usually followed by a main verb in the situative mood.' Below this, there is a table with columns for 'Left Cotext', 'Node', 'Right Cotext', and 'Source'. The table contains 10 rows of search results, each with a numbered entry, a snippet of text containing the word 'fela', the word itself, a snippet of text following it, and the source text ID.

	Left Cotext	Node	Right Cotext	Source
1	bbe gore koko ga a qwabele rena bana	fela	. Ge a se a tsoga gabotse, kok	Text4
2	e sa tsebe gore na o rata go no e etela	fela	, a e bone ke moka re boele gae	Text2
3	e sefateng sa gaChuene. Ge o etšwa	fela	ka sefate sa gaChuene, o tseni	Text3
4	bone. O ka se tsebe. Mpša e a goba	fela	ga re na taba nayo. Rena re a	Text1
5	i. Taba ke gore ga se ka bolela maaka	fela	, ke boletše maaka le nnete. Ki	Text3
6	na ge ke re ga se ka khora. Ke nyaka	fela	gore o tsebe gore ga se ka kho	Text2
7	ntshetla pelo. O a bona Sebego ga se	fela	moagišani wa ka eupša ebile ke	Text2
8	o. Ngwedi le wona ga o gona go setše	fela	dinaledi. Ga re na bothata bja	Text1
9	go nyama Matsepe o realo. " Ke swere	fela	diaparo ; borokgo le sekhipha.	Text1
10	šo! Motho wa go ba le motswadi o tee	fela	. ' O tsebe gore le basetsana b;	Text2

Figure 11: When working with POS-tagged corpora, an "Auto-search" automatically takes the part of speech into account

	Left Cotext	Node	Right Cotext	Source
1	morago ga go homola sebakanyana, "	ke	nyaka go ya go nyaka khehl	Text2
2	gago tša go robala ka ntlong ye kgolo -	ke	robotše ka ntlwaneng ya ka	Text1
3	hehla ya gešo e šoma Makgoweng . . .	ke	ra gore Gauteng. Ga ke tse	Text4
4	ga se tšhoše. Pele ke bolediša motho ,	ke	mo lebelela sefahlego pele. Text3	
5	seto. "Ke bolela ke emelela ka maoto ,	ke	thintha borokgo. "O ra gore	Text2
6	orwa ke bašemane. "Ga ke e tsebe. A	ke	re re tla no botšiša. "Ke a k	Text1
7	ilile?" "Re tla tseba gabotse monna. A	ke	re re tla botšiša batho? O h	Text1
8	e a tseba fao re yago monna Kholo. A	ke	re re a tseba gore re ya Leb	Text1
9	a le maseme. "Re a e tseba monna! A	ke	re ke a go botšiša gore e tla g	Text1
10	la go ya go nyaka khehla ya gešo. A	ke	re o tla nthuša, Kholo? "Set	Text2
11	nola mpa. "Bjalo o re nna ke direng? A	ke	re o a bona gore dijo di fedil	Text2
12	a go laela ga se ra kwana ka tšona. A	ke	re maabane re kwane gore	Text3
13	otšiša monna yola. "O tla mmotšiša a	ke	re? "Kgarebo o a mpotšiša. Text3	

Figure 12: A search for all instances of the subject concord *ke* only ("ke_SC"), and not of the copulative particle *ke*, agentive particle *ke*, or auxiliary verb *ke*

7. Conclusion

In this article it has been shown that the seamless integration of corpus query functionality within a dictionary compilation environment has now become a reality. The corpora may either be plain texts, or texts annotated for part of speech (POS). A method using minimal datasets was described with which such POS-tagged corpora may swiftly be obtained for resource-scarce languages. As a result, corpora can henceforth be queried more intelligently during dictionary compilation — and this for *any* language.

Acknowledgements

The additional research presented in this article was sponsored by *TshwaneDJe HLT*. A special word of thanks goes to Malcolm MacLeod who wrote most of the code that drives *TshwaneLex*'s corpus tool. Gilles-Maurice de Schryver would like to thank *Ghent University* for its continued support of his field trips to South Africa. Guy De Pauw is funded as a Postdoctoral Fellow of the *Research Foundation — Flanders (FWO)*.

Endnotes

1. The T-1 feature describes a single tag, while T+1 encodes all possible tags for the W+1 word. This is a consequence of the left-to-right tagging process. When disambiguating the focus

word, the right-hand side is not disambiguated yet, while the left-hand side is.

2. We also attempted the data-driven taggers MBT, TnT and MXPOST on this data set, but none of them were able to outperform MaxTag.
3. For unknown words there is a sudden decrease in tagging accuracy in the seventh partition, followed by a slow restoration of the linear learning curve. Error analysis indicates that this is due to the fact that the sixth partition contains a significant amount of idiosyncratic 'unknown words', disturbing the training phase when evaluating on the seventh partition.

References

- Berger, A.L., S. Della Pietra and V.J. Della Pietra.** 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics* 22(1): 39-71.
- Brants, T.** 2000. TnT — A Statistical Part-of-Speech Tagger. *Proceedings of the Sixth Conference on Applied Natural Language Processing (ANLP 2000), Seattle, USA*: 224-231.
- Daelemans, W., J. Zavrel, A. van den Bosch and K. van der Sloot.** 2003. *MBT: Memory Based Tagger, Version 2.0, Reference Guide*. Technical Report Series 03-13. Tilburg: ILK Research Group.
- De Pauw, G., G.-M. de Schryver and P.W. Wagacha.** 2006. Data-Driven Part-of-Speech Tagging of Kiswahili. Sojka, P., I. Kopecek and K. Pala (Eds.). 2006. *Text, Speech and Dialogue, 9th International Conference, TSD 2006, Brno, Czech Republic, September 11–15, 2006, Proceedings*: 197-204. Lecture Notes in Artificial Intelligence (LNAI), subseries of Lecture Notes in Computer Science (LNCS), volume nr. 4188. Berlin: Springer-Verlag.
- Giménez, J. and L. Màrquez.** 2004. SVMTool: A General POS Tagger Generator Based on Support Vector Machines. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal*: 43-46.
- Heid, U. et al.** 2007. *The IMS Textcorpora and Lexicon Group* [online]. Available: <http://www.ims.uni-stuttgart.de/projekte/corplex/>.
- Hurskainen, A.** 1999. SALAMA. Swahili Language Manager. *Nordic Journal of African Studies* 8(2): 139-157.
- Joffe, D. et al.** 2007. *TshwaneLex Suite* [online]. Available: <http://tshwanedje.com/tshwanelex/>.
- Kilgarriff, A. et al.** 2007. *Sketch Engine* [online]. Available: <http://www.sketchengine.co.uk/>.
- Kupiec, J.** 1992. Robust Part-of-Speech Tagging Using a Hidden Markov Model. *Computer Speech and Language* 6: 225-242.
- Le, Z.** 2004. *Maximum Entropy Modeling Toolkit for Python and C++* (Technical Report) [online]. Available: http://homepages.inf.ed.ac.uk/s0450736/maxent_toolkit.html.
- Marcus, M., B. Santorini and M. Marcinkiewicz.** 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics* 19(2): 313-330.
- Mojela, M.V. et al.** 2007. *Pukuntšuthaloši ya Sesotho sa Leboa ka Inthanete* (Explanatory Sesotho sa Leboa Dictionary on the Internet) [online]. Available: <http://africanlanguages.com/psl/>.
- Ratnaparkhi, A.** 1996. A Maximum Entropy Model for Part-of-Speech Tagging. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Somerset, USA*: 133-142.
- Scott, M.** 2007. *WordSmith Tools* [online]. Available: <http://www.lexically.net/wordsmith/>.
- Van Halteren, H., J. Zavrel and W. Daelemans.** 2001. Improving Accuracy in Word Class Tagging through Combination of Machine Learning Systems. *Computational Linguistics* 27(2): 199-230.

Addendum 1: Part-of-speech tags used for the Northern Sotho tagging experiment described in Section 5

POS-tag	Meaning	Example
ADJ	adjective	... nako tše dintši o a tseba.
ADV	adverb	Gape kgekolo yela yona ...
AGTp	agentive particle ('ke')	Wena o dirwa ke gore o kwana ...
ASP	aspectual prefix	Ba be ba sa hwile ka boroko.
ASP+go	aspectual prefix + relative marker ('go')	... tšogo ...
AUX_V	auxiliary verb	Ba be ba sa hwile ka boroko ...
AUX_V+go	auxiliary verb + relative marker ('go')	Ga go na yo a ilego a bona ...
CONJ	conjunction	"Ga se ba go kwa ge o etšwa?"
CONp	connective particle ('le')	Ke realo le nna ke bolela ka ...
COP_V	copulative verb	Ga go na yo a nkwelego ge ke ...
COP_V+go	copulative verb + relative marker ('go' OR 'ng')	... yona ye e lego kgauswi ...
COPp	copulative particle ('ke', 'ga se')	Taba ke gore ke di boletše ...
CP15	infinitive prefix cl. 15 ('go')	... tšeo tša gago tša go robala ...
DEM	demonstrative	Gape kgekolo yela yona e a ...
DEM-COP	demonstrative copulative	Yena šo!
ENUM	enumerative	... wa di bolela ka nako e tee .
EXCL	exclamation	O reng o ntšhoša ngwana tena?
FUT	future morpheme ('tlo', 'tla')	... re tla ngwathagana tše tša ...
FUT+go	future morpheme ('tlo') + relative marker ('go')	... tlogo ...
HRTp	hortative particle	A re goge.
IDEO	ideophone	Ke no re ge ke re phaphara ...
INSp	instrumental particle ('ka')	... o mpotšiša ka sehebehebe.
INTERJ	interjection	Ai!
ka+eng	instrumental particle ('ka') + question word ('eng')	O tla nkiša kang ka gore le ...
ke+eng	copulative particle ('ke') + question word ('eng')	Molato keng?
ke+ka	first person singular ('ke') + potential morpheme ('ka')	Ke bona nka kgona go thuša ...
LOCp	locative particle	... ke robotše ka ntlwaneng ya ...
MPG	merged particle group	... fela ga re na taba nayo .
N	noun	... tlhokomelo e le ka lebaka la ...
N_PROP	proper name	Seboko o mpotšiša ka ...
N+ng	noun + locative marker ('ng')	Ke realo mafelelong ke ...
NEG	negative morpheme ('ga', 'sa', 'se', 'ga se')	Ga se ba go kwa ge o etšwa?
NEUT	neutral subject marker ('e')	E sa le gosasa.
OC	object concord	"Ga se ba go kwa ge o etšwa?"
OC1sg+V	object concord 1p sg + verb	Bjale o nyaka go ntlhanogela?
OCcl1+V	object concord cl. 1 + verb	Ke a mmotšiša ke sa na le ...
PAST	past tense morpheme ('a')	... ba a phetha modiro wa bona.
PC	possessive concord	... sa ka godimo ga letsogo ...
POSS	possessive	Diaparo tšagwe di hlwekile.
POT	potential morpheme ('ka')	... ga go na seo o ka se dirago ...
POT+go	potential morpheme ('ka') + relative marker ('go')	Go tla ba le kago ya mašole ...
PRES	present tense morpheme ('a')	... kgekolo yela yona e a kwa.
PRO_abs	absolute pronoun	Ke realo le nna ke bolela ka ...

POS-tag	Meaning	Example
PRO_comm _poss	communal possessive pronoun	... sa khehla ya geno ?
PRO_poss	possessive pronoun	... seatla sa ka godimo ga ...
PRO_quant	quantitative pronoun	... re aparetšwe ke seetša gohle .
Punc	punctuation	Wena o re ke swereng ka mo?
QSTp	question particle	... le eng na ?
SC	subject concord	" Ke šikinya hlogo."
SC_ind	indefinite subject concord ('go')	... ka fao go ra gore ga se ba ...
TMPp	temporal particle ('ka')	... tswalelwa ka Labohlano.
unknown	unknown POS	
V	verb	"Ga se ba go kwa ge o etšwa ?"
V+eng	verb + noun ('eng')	Wena o re ke swereng ka mo?
V+go	verb + relative marker ('go' OR 'ng')	... ke o swerego ka seatla sa ...
V+ng	verb + plural marker ('ng')	... a sa tsebe gore o bolelang ...

Is Lexicography Making Progress? On Dictionary Use and Language Learners' Needs

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

Abstract: This article sets out to explore the ways native speakers as well as foreign language learners use dictionaries and the strategies dictionary users adopt in the language acquisition process. The basis for this article is a corpus of six books (in chronological order Atkins (Ed.) 1998, Nesi 2000, Tono 2001, Humblé 2001, Sin-wai (Ed.) 2004 and Thumb 2004) that look at both the usage of bilingual, monolingual and bilingualised dictionaries and the users' behaviour in the consultation process. Both the bilingual and monolingual dictionaries seem to be used independently, depending on whether the user wants to utilise them for comprehension, translation or production with regard to a foreign language. As pointed out in the literature on lexicography, some of these dictionaries, though they have undergone many changes over the years, still have serious limitations as learning tools, but the user's performance is also under investigation in empirical research, with the aim of optimising dictionary effectiveness as well as developing the language learner's skills.

Keywords: LEXICOGRAPHY, MONOLINGUAL, BILINGUAL, BILINGUALISED, DICTIONARIES, LANGUAGE ACQUISITION, DICTIONARY USER, LOOK-UP STRATEGIES, THINK-ALoud, TRANSLATION, CORPUS, SKILLS

Opsomming: Maak leksikografie vordering? Oor woordeboekgebruik en taalaanleerders se behoeftes. Die doel van hierdie artikel is om die maniere te verken waarop beide moedertaalsprekers en vreemdetaalaanleerders woordeboeke gebruik en die strategieë wat woordeboekgebruikers toepas by die taalverwerwingsproses. Die basis vir hierdie artikel is 'n korpus van ses boeke (in chronologiese volgorde Atkins (Red.) 1998, Nesi 2000, Tono 2001, Humblé 2001, Sin-wai (Red.) 2004 en Thumb 2004) wat kyk na sowel die gebruik van tweetalige, eentalige en vertweetalige woordeboeke as die gebruikers se gedrag by die raadplegingsproses. Beide die tweetalige en eentalige woordeboeke word skynbaar onafhanklik gebruik, afhangende daarvan of die gebruiker hulle wil aanwend vir begrip, vertaling of produksie met betrekking tot 'n vreemde taal. Soos die literatuur oor leksikografie aantoon, het sommige van hierdie woordeboeke, alhoewel hulle deur die jare baie veranderinge ondergaan het, steeds ernstige beperkinge as leerwerktuie, maar die gebruiker se prestasie word ook in empiriese navorsing ondersoek met die doel om sowel die effektiwiteit van woordeboeke te verhoog as om die taalaanleerder se vaardighede te ontwikkel.

Slutelwoorde: LEKSIKOGRAFIE, EENTALIGE, TWEETALIGE, VERTWEETALIGDE, WOORDEBOEKE, TAALVERWERWING, WOORDEBOEKGEBRUIKER, NASLAANSTRATEGIEË, HARDOP-DINK, VERTALING, KORPUS, VAARDIGHEDE

1. Books under review

Although some results given in the six books under discussion may now appear outdated, their evaluation of earlier dictionary-use research and its methodologies represents a mass of information on the progress made in lexicography these past decades. The authors who contributed to the volume edited by Atkins (1998) focus more particularly on language learners' and student translators' look-up strategies. They investigate specific questions relating to the skills users should possess to gain the most from dictionaries and improvements to dictionaries to suit the users' expectations and needs. In these researches, a number of tests and linguistic exercises are included to assess whether user-oriented dictionaries meet the students' expectations and needs. Nesi (2000) presents a series of five experiments for adult English learners and the part played by dictionaries in reading comprehension. Some consultation problems and selection of wrong entries may cause serious misinterpretations. Tono (2001) is a very comprehensive book, giving an extensive review of research findings from the 1980s to the 2000s on dictionary use in reading and vocabulary acquisition. It broaches a discussion of how to evaluate users' dictionary abilities and look-up efficiency by means of skills tests. Mainly English–Japanese bilingual dictionaries are the focus of this research. Humblé (2001) reviews literature on dictionary use and examines different dictionaries in order to elaborate his vision of a new kind of dictionary which would maximize users' vocabulary. Sin-wai's edited collection of papers (2004) looks at the use of bilingual dictionaries for translation and new IT technology in this domain. The emphasis of Thumb's (2004) study, finally, is on English–Chinese bilingualised dictionaries and Hong-Kong students' look-up behaviour and strategies.

This very valuable sum of information about experimental methods will contribute to broaden readers' background knowledge and provide them with a basis for future research. A review of these major studies from well-known specialists in linguistics, translation and lexicology, offers an opportunity to examine users' behaviour in the lexical acquisition process of dictionary consultation, to address questions on data collection, on the strengths and weaknesses of the authors' findings, and to look ahead to the development of lexicography in the 21st century.

2. Defining the dictionary user

As shown by these studies, what makes the compiling of a dictionary complex and problematic is that there is not a prototypical user, but a whole variety of users, with different cultural backgrounds which may affect their abilities. All articles are subject to possible criticism as regards the sample of users targeted or the methodology used. Other external factors such as the difficulty of the task, the time allocated, the type of dictionary used, and the frequency of con-

sultation, if not taken into account when observing subjects' behaviour, may invalidate some of the findings. However, authors are generally aware of possible fallacies in the results of their investigation.

Dictionaries can prove hard to compare as they use different approaches and target divergent groups of prospective users. Indeed, there are, among other categories, subgroups of users: English native speakers, non-native speakers of English such as Chinese or Japanese students, language learners of many different nationalities studying a foreign language, or translators (Tono 2001: 114). As there is a great diversity of potential users, the situations in which dictionaries are used are as varied. Dictionaries can be consulted in language teaching environments such as in formal circumstances in a classroom, in a professional situation for translators or in the more informal situation of a visit to a foreign country. The tendency of researchers to narrow their studies of dictionary use to the language-learning context (more easily observable in a 'controlled' experimental environment) has at times its limitations, because the needs of a subgroup of learners, but not of all potential learners are taken into consideration.

Studies of the way language learners use their dictionaries provide valuable information for both lexicographers and language teachers. They aim at improving not only the dictionaries but also the look-up abilities of their users. As Humblé (2001: 21) notes, the best dictionary compiler is in fact one who is him-/herself a dictionary user and a language learner. In fact, the points of view of experienced language teachers are more unbiased and reliable than those of some lexicographers and academic researchers whose knowledge of dictionary use and language teaching can be limited. It is true that existing dictionaries tend to be made by native speakers who, judging on their intuition rather than on empirical knowledge, are less aware of the language difficulties of their mother tongue than language learners (Humblé 2001: 73). Humblé combines, as he describes himself, the expertise of a foreign language learner, a university lecturer, a researcher, and a dictionary compiler. His exposure to lexicography and language teaching is very valuable in assessing how current dictionaries fail in conveying information and how they can be improved.

The use of a dictionary has now become an indispensable component of home and academic life. Three main types of dictionaries: monolingual, bilingual and bilingualised dictionaries, and two main categories of users: native speakers and foreign language learners can be identified. It must be stressed that some cultures are more dictionary literate than others. In France, for instance, French speakers learning English have a range of monolingual dictionaries to hand, but the two most popular are perhaps *Le Petit Larousse* and *Le Petit Robert*. As the latter is more literary than the former, users will tend to utilise it for academic research and translation. *Le Petit Larousse* targets a wider audience, not necessarily students or academics: it is more compact and, unlike *Le Petit Robert*, provides numerous illustrations. Nevertheless it can be said that the choice of a dictionary is quite arbitrary, often a question of taste. Some grow up with a particular dictionary and, being familiar with its layout and conventions, tend to keep using it.

Monolingual dictionaries of the type of *Le Petit Larousse* and *Le Petit Robert* are no longer language 'tools' for foreign learners but rather 'repositories' (Humblé 2001: 24) whose taxonomy, in the tradition of the various editions of the *Dictionnaire de l'Académie*, originally endeavours to preserve standard language and proscribe non-standard usage. Nesi (2000: 102) has proven that language background and culture somehow have an influence on the way a user consults a dictionary and on his/her look-up efficiency. For English-speaking societies, the proximity with English (used as a lingua franca) is certainly an asset, while for cultures whose mother tongue is remote from English, the need for contextualised English dictionaries is paramount to help them understand the cultural background with which they are unfamiliar.

In the 1970s, the traditional monolingual dictionaries gave birth to a new concept: a descriptive rather than a normative language learner's dictionary. The main disadvantage of a learner's dictionary is that by simplifying its definitions and reducing the number of entries to make it accessible to students not familiar with the language, it tends to exclude 'the least frequent words that cause most trouble to native speakers' (Humblé 2001: 35). It might also restrict it to the least proficient language learners, as advanced learners look for a more exhaustive number of entries and less frequent words in the linguistic repertoire. A study conducted by Nesi on the two editions of a learner's dictionary, COBUILD 1 and COBUILD 2, shows 'entries are getting longer, and the number of examples within each entry is increasing' (2000: 137) to suit a greater diversity of users. Although the size of some of these entries makes it denser and therefore more complex for the user, especially beginners, larger and more comprehensive dictionaries are highly rated by reviewers.

3.1 Monolingual versus bilingual: which is the better?

One of the difficulties dictionary users encounter in using a monolingual dictionary is the abstruse 'metalanguage of the definitions' (Thumb: 2004: 4) which they find too lengthy or difficult to decode. Sometimes, they need to look up several words contained within a definition. This inability to understand definitions leads to frustration or inhibition which prevents language acquisition and causes the avoidance of monolingual dictionaries, whatever their linguistic skills in L2 and the task to be performed (Atkins and Varantola in Atkins 1998: 38). Learners are also frequently misled and confused by multi-sense entries,¹ being unable to retrieve the appropriate meaning of a word from all its possible meanings. Likewise, attempts to find information on collocations and the grammar of a particular structure often ends up in failure. Tracing an idiomatic phrase or realising that a phrase is an idiom are also problematic for many dictionary users.

Most studies conducted show that foreign language learners normally favour bilingual to monolingual dictionaries (Thumb 2004: 5), Nesi, however, adding that they generally 'use them badly' (1999: 65). Similar findings, that

both language learners and translators favour bilingual dictionaries, emerge in Atkins (1998). For translation, the research conducted by Bogaards (see Nesi 2000: 22) states that bilingual dictionaries are more effective, although they do not favour vocabulary retention. In fact, Tomaszczyk (see Nesi 2000: 6) showed that monolingual dictionaries are more useful to foreign dictionary users who are advanced learners. They are said to improve their vocabulary and help them acquire sophistication. High proficiency learners, unlike low proficiency learners, often take to monolingual dictionaries when bilingual dictionaries do not suffice for finding extra information on a word meaning. Figures show that monolingual dictionary consultation increases with learners' linguistic maturity (Nesi 2000: 29).

When a dictionary is consulted to do a translation, figures demonstrate that bilingual dictionaries are overall more effective than monolingual dictionaries, resulting in a higher success rate (Atkins and Varantola in Atkins 1998: 99). The most effective translations are produced by those who consulted a dictionary prior to translating. Those translating into their own language appear to be more satisfied with their translation because of their familiarity with the language, but this does not necessarily imply a successful translation. Trained translators in their search for perfection do not use one but several dictionaries. They consult a variety of sources (a whole range of reference books and text bases) to corroborate their results.

Advanced English learners use dictionaries more than elementary learners (Thumb 2004: 11). At an early stage of their language training, foreign language students have limited communication skills and vocabulary abilities. They are confronted with simpler messages which do necessitate the frequent use of a dictionary. At this level, the dictionary needs to display clear and learner-friendly examples. As they become more proficient and tend to native-like fluency, learners wish, on the one hand, to make use of a more elaborate lexis with the use of metaphors, culture-specific idioms and correct collocations and, on the other hand, to develop knowledge of styles and 'intellectualise' (Castro in Sin-wai 2004) the culture of the target language.

Can the profile of the successful dictionary user be established? It is obviously something of a myth to believe that there will be an 'ideal' user in the same way that the concept of a 'perfect' dictionary is purely ideological. However, Thumb (2004: 105) provides some answers. The user should already have a high level of language proficiency, use both L1 (mother tongue) and L2 (foreign language) sections of the dictionary and be familiar with dictionary conventions. In other words, practice certainly makes the user, if not perfect, at least better skilled.

3.2 Bilingualised dictionaries

Bilingualised dictionaries bring the synergy of two dictionaries in one with exposure to L2. They are hybrids of the monolingual and bilingual dictionary.

In a bilingualised dictionary, each headword entry is followed by a definition in L2 and an L1 translation. Studies show that users of a bilingualised dictionary have a low failure rate in word retrieval (Laufer and Hadar 1997: 192). The translation sought in such a dictionary has a reassuring effect, for the non-native user finds confirmation for his/her understanding of the meaning of the target word and for making the right decision. Dídac et al. (2006) argue, however, that in print bilingualised dictionaries, the user does not necessarily have exposure to L2, as some pass over the monolingual L2 part and read only the L1 translation, losing the potential offered by the definition. Newly released electronic bilingualised dictionaries offer great promise for the future. Some display the monolingual L2 text and defer the L1 translation so that users do not encounter it immediately. Others, such as the *Ultralingua Dictionary*, separate both definition and translation with several hyperlinks, which enable users to choose their language and encourage them to consult both monolingual and bilingual sections. Another disadvantage of the print bilingualised dictionary is that it provides solely an L2 + L1 section but not the reverse. Deferred bilingualised dictionaries keep every section separated, but at the same time inter-related by means of interactive menus.

4. Users' and dictionary strategies

Users' lack of knowledge of dictionary conventions is often regarded as one of the main causes for their look-up failures. Lexicographical researches have been conducted with the aim of improving the dictionary as well as improving the look-up abilities of the users. By studying the way users consult their dictionaries, researchers can identify strategies, those that fail and those that assist users in being more efficient in their search. As Nesi (2000: 11) points out, questionnaires to evaluate users' strategies, speed and performance in dictionary consultation have been subject to controversy. Indeed, it is rather difficult to measure whether the respondents' answers are actually a reflection of what they do or think or whether they are putting forward a positive view of their abilities, embellishing what they actually do and think. Research based on observation of dictionary users' behaviour can prove holistic in a contrived experimental situation. The size of the group to be investigated can also be a major disadvantage. A group of less than 20 participants does not seem enough to constitute a representative sample, but the larger the number, the more difficult it is to monitor.

It is necessary to be aware of users' different motivations for consulting a dictionary when evaluating the best possible strategies for successful look-ups. A linguistic activity of some sort such as speaking, reading and translating is what triggers the use of a dictionary. Tarp (in Sin-wai 2004: 28) contrasts two types of situations where the user will consult a dictionary: 'communication-orientated' and 'knowledge-orientated' situations. Generally users' look-ups for new words in a dictionary are in decreasing order of preference for meaning,

pronunciation and spelling. Proficient users would look most specifically for typical collocations and idioms.

The less proficient in English an EFL student is the more he/she will need to resort to dictionaries in order to broaden his/her vocabulary. With language proficiency he/she also acquires a better judgement on when to consult a dictionary. His/her failure in using dictionaries might be due to a lack of expertise in dictionary use, as researchers generally notice. It has been found that 80% of the French speakers investigated for the EURALEX/AILA Research Project had no expertise in the use of a dictionary (Atkins and Varantola in Atkins 1998)². The foreign language learner's lack of dictionary-using skills has been widely observed (Nesi 1999), testifying to the little importance given to dictionary training in language teaching. Language users also often ignore the wealth of information that the dictionary contains and make little use of the extra features its microstructure has to offer (Tono 2001: 114). In addition, learners have problems in becoming sensitised to the register of words they are using. They may use colloquial vocabulary in formal situations and rather stilted vocabulary in casual contexts. In dictionaries, vocabulary is stylistically labelled to indicate technicality or level of informality.

Studies have been conducted to establish whether definitions or examples or both in dictionary entries contribute to a better understanding of the target word. Findings unsurprisingly tend to show that for comprehension 'new words were best learned when they were both defined and illustrated' (Tono 2001: 23). Examples are used to corroborate dictionary definitions and, by conveying linguistic information, they certainly enhance the learning process but it still needs to be determined whether this helps vocabulary retention. The question of whether authentic examples are more beneficial to language use than made-up ones is the source of some debate (Humblé 2001: 78, Nesi 2000). In the 1980s, corpus-based dictionaries started including examples from newspaper articles, and speeches recorded in naturally-occurring situations³. This concept of authenticity is, as Humblé points out, purely arbitrary: 'the fact that a native speaker produced an utterance does not render it authentic' (2001: 78). Made-up examples are, on the contrary, grammatically correct but not genuine, for the lexicographer would tend to combine several grammatical problems in one. The hackneyed example given for a long time in English grammars of French as illustrative of the possessive case, 'la plume de ma tante', sounds rather artificial out of context, and it is very unlikely a user will come across such an example in real life. Made-up examples have the benefit of illustrating a particular grammatical point or even combining difficulties, but there is no doubt that the use of authentic material has a psychological impact on users who feel they are exposed to the language as it is used.

To collect data on users' look-up strategies, think-aloud is widely accepted to be an effective tool (Thumb 2004: 33) for evaluating the thinking process of a dictionary user during consultation. The process of dictionary consultation implies 'a series of deliberate, metacognitive and cognitive steps learners take in

acquiring and retrieving dictionary information' (Thumb 2004: 85). It has been used in fields such as psychology and science to assess knowledge acquisition. Thumb has studied the look-up strategies of a sample of advanced learners of English in Hong Kong while consulting a bilingualised English–Chinese learner's dictionary. It has been argued that some look-up strategies will enable the user to be successful in retrieving the appropriate word, while others will be misleading (Thumb 2004: 32). Her analysis of think-aloud data has enabled her to identify seven different dictionary-use strategies⁴:

- The *ignoring* strategy: The subject looking up both monosemous and polysemous words does not make use of the full ancillary information (such as the grammatical constraints, word class, etymology, etc.) available in an entry.
- The *assuming* strategy: The subject looking up monosemous words uses ancillary information, but assumes that the meaning in an entry is appropriate and does not apply it to the context in which it is used.
- The *minimizing* strategy: The subject looking up polysemous words minimises the use of definitional and ancillary information when searching for a word by reading just the first or first few senses of a word.
- The *checking* strategy: The subject infers the possible meaning of a word by checking if it matches with the dictionary definition.
- The *paraphrasing* strategy: The subject just rewords the chosen definition of a word.
- The *stretching* strategy: The subject infers a context-fitting meaning for the target word common to all definitions in an entry.
- The *maximising* strategy: The subject scans all the meanings of the target word in an entry to reduce it to one.

Thumb's results indicate that both higher and lower proficiency subjects used the L1 as frequently as the L2 part of bilingualised dictionaries, but would prefer L1 for some words and L2 for others (Thumb 2004: 98). Using different look-up strategies does not guarantee success, but learners using the *assuming* and *checking* strategies proved on the whole more successful in retrieving the correct meaning of a target word.

5. Dictionaries and translation

The translation activity and users' strategies have been the focus of Thumb's analysis (2004) in the context of Hong Kong students using a bilingualised dictionary for reading comprehension. Sin-wai's book (2004), most particularly focusing on Chinese/English translation, gives a complementary survey of the progress made by lexicography in the field of translation. This is a topic on

which little research has been conducted. The book brings together experts in translation studies and translation practice and lexicographers, mostly of Chinese origin. Not only do they look at new trends in dictionary compilation, but they also discuss the translation friendliness of monolingual, bilingual and ultimately bilingualised dictionaries. The bilingual dictionary is often inadequate, unless words are organised and cross-referenced, in the same way as in thesauri, according to semantic domain templates, as suggested by Moe (in Sin-wai 2004: 55-62). Indexed words can thus be compared and contrasted, making it easier for the user to find an equivalent in the entire lexicon. The development of new technologies for the translation market (bilingual corpora, computer translation software and machine translation) has provided quickly and easily updated resources. Translation memory enables the storage of a database of previously translated sentences which could be provided to the user at any time (Yihua in Sin-wai 2004: 104). In our globalized world with English expanding in China, Kam-mei and McArthur (in Sin-wai 2004: 126) imagine that tomorrow's dictionary for the Hong Kong population will be biliterate and trilingual in accordance with the language policy promoted by the government. Other issues at stake in the book are the translation of metaphors and other culture-bound references. No doubt, as Yanchun and Jianhua (in Sin-wai 2004: 177-185) make it clear, translation is somewhat of a 'paradox'. Translation cannot just be a mechanical word-to-word translation, as no complete equivalents exist in two different languages. Words differ on account of their connotations and denotations. In the process of translation, something is inevitably lost and the translation of the source language will often need to compensate. In the next decade, with the improvement of electronic dictionaries, major advances can be expected in their application to translation.

6. Producing the dictionary of the future

Most books discussed in this article make observations in their concluding chapters on future trends in lexicography, speculating on what the dictionaries of tomorrow will look like. Nesi (2000) notes that print dictionaries are growing in size with the risk of overloading the users with too much information. As a bigger dictionary is not easy to carry around and the information is more difficult to retrieve, the solution advocated is to become electronic. Print dictionaries are constrained by space, while electronic dictionaries have no size limitations. The dictionary of the future, as seen by Rogers and Ahmad (in Atkins 1998: 193), will enable users 'to draw on electronically-stored data-bases of terms with semantically-relevant navigational paths [and] will support them in creating their own data-base'. Humblé (2001: 97) substantiates the claim that the ideal dictionary should be electronic, stressing the need to combine in a dictionary the tools for decoding with the largest number of lexical items and for encoding with the most detailed guidance on usage, syntactic features and collocates: 'A Spanish-English dictionary is supposed to help Spanish speakers

with the understanding and production of English while at the same time assisting English speakers with the understanding and production of Spanish' (Humlé 2001: 84).

With no space constraints, the electronic dictionary⁵ can give access to a diversity of sources from the World Wide Web with cross-referenced links to synonyms, antonyms and thesauri. It can be multifunctional, including as much information as possible on meaning, etymology, grammar and collocations. Other multimedia applications include the use of sound-like spellings, attractive animation and visual elements such as photographs and video clips which, on account of their game-like approach, favour vocabulary acquisition. These appeal to users who find access to information not only interactive and enjoyable, but also much faster than consulting a print dictionary. With a concordancer and software such as WordSmith and Multiconcord, students also have the possibility to compile their own 'virtual corpus' (Varantola in Atkins 1998: 190). These corpus data allow the study of vocabulary and collocations, particularly syntactic and lexical units, in texts extracted from Internet sources. Corpora also provide a wealth of immediately accessible examples with which no paper dictionary can compete.

7. Conclusion

The research conducted in the books under review is sometimes inconclusive and the statistical validity of the samples and the findings can be debated. They raise many questions, not always enlightening the reader on how to answer them. The process of dictionary consultation showing the strategies of users is a very complex phenomenon to observe. Tono (2001: 84) remarks with scepticism that 'not a single reliable measurement of dictionary using skills has become available so far'. Results to record subjects' behaviour are often difficult to interpret and seem artificial in a non-natural environment. It would be advisable, as he (Tono 2001: 36) suggests, to have them verified 'in a more natural classroom setting' or in real-life social situations, both formal and informal. Other findings prove statistically to be rather obvious or common-sense (the use of a dictionary affects the amount of time spent to complete a test), but they open lines of investigation on how to improve both the users' skills and the general design of the dictionary. Future researchers learning from their predecessors will certainly be guided to prevent the same pitfalls. For an improved bilingual dictionary, researchers of both L1 and L2 are now expected to collaborate. In the 1930s, Ščerba had already imagined a reference tool to assist users in decoding and encoding, combining four dictionaries in one and providing translation 'A→B and B→A for users with the mother tongue A and A→B and B→A for users with the mother tongue B' (Berkov 1996, quoted by Humlé 2001: 98). This is hardly conceivable in paper-based form. However, in the near future, electronic dictionaries or CD-ROM versions will be seen becoming multifunctional, combining in one programme the potential of the

monolingual, bilingual and bilingualised dictionaries. CD-ROMs now provide users with instructions and a virtual guided tour of their dictionary potential, ensuring participants some form of training in dictionary skills they usually lack.

Tono (2001: 216-217) is confident that new IT technology, which creates 'a user-friendly learning environment' will have some influence on subjects' behaviour and will improve their dictionary skills, by giving them access, in the most interactive way, to large lexical databases. Facing the computer, students do not experience the inhibitive fear and panic which are often prevalent in the classroom situation. Despite general enthusiasm towards on-line computer-based dictionaries and CD-ROMs, however, little research has been conducted on their efficiency. At this stage, it is hard to assess whether fast searches in an electronic dictionary actually improves vocabulary acquisition and permits retention, but the electronic dictionary of the third millennium holds many promises for better skilled and more proficient foreign language speakers.

Notes

1. A single entry can contain polysemous meanings ranging from general to more technical fields. OALECD for instance gives four main senses for the word 'blade' (Thumb 2004: 96).
2. Pupils in France are normally taught to use a dictionary as early as primary school: at an early age (in CE1), children's picture dictionaries and then gradually (from CM1 onwards), *Le Petit Larousse*.
3. Moon (1989) argues that the selection of texts used to compile corpus-based dictionaries are ideologically 'dominated by published texts written by middle-class writers for a middle-class readership' (quoted by Benson in Sin-wai 2004: 46).
4. Nesi (2000: 91) suggests, moreover, that confusion with 'some phonologically and orthographically similar word' is a major cause of error.
5. Electronic dictionaries fall into three categories: 'hand held devices, dictionaries on disk and CD-ROM, and dictionary Websites' (Nesi 2000: 142).

References

- Atkins, B.T. Sue (Ed.). 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer.
- Berkov, Valerij. 1996. Passive vs. Active Dictionary. A Revision. *Euralex '96 Proceedings. Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 547-550. Göteborg: Göteborg University, Department of Swedish.
- Bogaards, Paul. 2001. Dictionnaire pédagogiques et apprentissage du vocabulaire. *Cahiers de Lexicologie* 59: 93-107.
- COBUILD 1: Sinclair, John M. (Ed.). 1987. *Collins COBUILD English Dictionary*. London/Glasgow: Collins.

- COBUILD 2: Sinclair, John M. (Ed.). 1995. *Collins COBUILD English Dictionary*. London: Harper-Collins.
- Dídac, Pujol, Corrius Montse and Joan Masnou.** 2006. Print Deferred Bilingualised Dictionaries and Their Implications for Effective Language Learning: A New Approach to Pedagogical Lexicography. *International Journal of Lexicography* 19(2): 197-215.
- Humblé, Philippe.** 2001. *Dictionaries and Language Learners*. Frankfurt: Haag und Herchen.
- Lafer, Batia and L. Hadar.** 1997. Assessing the Effectiveness of Monolingual, Bilingual, and 'Bilingualised' Dictionaries in the Comprehension and Production of New Words. *The Modern Language Journal* 81(2): 189-196.
- Moon, Rosamund.** 1989. Objective or Objectionable? Ideological Aspects of Dictionaries. *English Language Research* 3: 59-94.
- Nesi, Hilary.** 1999. The Specification of Dictionary Reference Skills in Higher Education. Hartmann, R.R.K. (Ed.). *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Sub-Project 9: Dictionaries*: 53-66. Berlin: Free University Berlin.
- Nesi, Hilary.** 2000. *The Use and Abuse of Learners' Dictionaries*. Lexicographica. Series Maior 98. Tübingen: Max Niemeyer.
- OALECD: Hornby, A.S. and P.T. Lee (Eds.). 1994. *Oxford Advanced Learner's English-Chinese Dictionary*. Hong Kong: Oxford University Press.
- Sin-wai, Chan (Ed.).** 2004. *Translation and Bilingual Dictionaries*. Tübingen: Max Niemeyer.
- Thumb, Jenny.** 2004. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary: A Think-Aloud Study*. Tübingen: Max Niemeyer.
- Tomaszczyk, I.** 1979. Dictionaries: Users and Uses. *Glottodidactica* 12: 103-119.
- Tono, Yukio.** 2001. *Research on Dictionary Use in the Context of Foreign Language Learning*. Tübingen: Max Niemeyer.
- Ultralingua Dictionary*. <http://www.ultralingua.net>.

Doen die *Woordeboek van die Afrikaanse Taal Deel XII* dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie?

Ilse Feinauer, *Departement Afrikaans en Nederlands, Universiteit van Stellenbosch, Stellenbosch, Republiek van Suid-Afrika (aef@sun.ac.za)*

Opsomming: Hierdie resensieartikel wil vasstel in watter mate WAT deel XII aan die missie van die WAT voldoen, naamlik "om die Afrikaanse taal en die gebruikers daarvan te dien, asook die leksikografie in die algemeen, veral die Suid-Afrikaanse leksikografie". Daar word ondersoek ingestel na die voortekste, agtertekste en die sentrale teks, en aangetoon dat albei buitetekste geïntegreer is met die sentrale teks en in daardie opsig dus wel die leksikografie dien. Wat die sentrale teks betref, word tot die gevolgtrekking gekom dat die gebruikers van die Afrikaanse taal sowel as die Suid-Afrikaanse leksikografie meer daarby sal baat indien dit deegliker gekonseptualiseer en beplan sou word aan die hand van leksikografiese funksies. Die keuse van leksikale items wat die omvattendheid van die Afrikaanse leksikon illustreer en verteenwoordig sal ook aan die hand van 'n goed beplande korpus herbedink moet word. Frekwent gebruikte woorde sal dus nie langer as lemmatekens oor die hoof gesien word nie en die lengte van artikelinskrywings behoort ook meer gelykvormig gebalanseer te word. Sodoende sal die Afrikaanse taal en die gebruikers daarvan beter gedien word. Die Suid-Afrikaanse leksikografie sal daarby baat indien die WAT sou terugkeer na wat van 'n omvattende woordeboek verwag word, naamlik 'n meer gebalanseerde verteenwoordiging van makro- en mikrostruktuur. Dit behoort ook te lei tot 'n meer eweredige verspreiding van artikelomvang en minder onbewerkte lemmatekens.

Sleutelwoorde: AGTERTEKSTE, BUITETEKSTE, LEKSIKOGRAFIE, LEKSIKOGRAFIESE FUNKSIE, LEMMATEKEN, MAKROSTRUKTUUR, MEDIOSTRUKTUUR, MIKROSTRUKTUUR, OMVATTENDE WOORDEBOEK, ONBEWERKTE LEMMATEKEN, RAAMSTRUKTUUR, SENTRALE TEKS, VOORTEKSTE

Abstract: *Does the Woordeboek van die Afrikaanse Taal Part XII do it for the Afrikaans Language and South African Lexicography?* This review article seeks to discover the extent to which WAT Part XII complies with its mission, that is "to serve the Afrikaans language and its users, as well as lexicography in general, especially South African lexicography". The front matter, back matter and central list are investigated and it is shown that both outer texts are integrated with the central list and in that respect lexicography is therefore served. With regard to the central list, it is concluded that users of Afrikaans as well as South African lexicography will benefit more if this would be conceptualised more systematically and planned according to lexicographic functions. The choice of lexical items which illustrate and represent the comprehensivity of the Afrikaans lexicon would also have to be rethought on the basis of a well-

designed corpus. Frequently used words would then no longer be overlooked as lemma signs and the length of dictionary articles should also become more evenly balanced. In this way the Afrikaans language and its users will be served better. South African lexicography would benefit more if WAT were to return to what is expected of a comprehensive dictionary, namely a more balanced representation of the macro- and microstructure. This should also lead to less disparity in article size and fewer untreated lemma signs.

Keywords: BACK MATTER, CENTRAL LIST, COMPREHENSIVE DICTIONARY, FRAME STRUCTURE, FRONT MATTER, LEXICOGRAPHY, LEXICOGRAPHIC FUNCTION, LEMMA SIGN, MACROSTRUCTURE, MADIOSTRUCTURE, MICROSTRUCTURE, OUTER TEXTS, UNTREATED LEMMA SIGN

Inleiding

Met die verskyning van WAT XII word die letters P en Q voltooi. Deel XII beslaan 760 bladsye met 8 936 verklaarde lemmatekens waarvan 8 873 aan P gewy word en die oorblywende 63 aan Q. Hierdie deel verskyn in die 79ste bestaansjaar van die WAT en in die 80ste verjaardagvieringe in 2006 speel dit 'n sentrale rol. Dit is die eerste deel onder die hoofredakteur dr. W.F. Botha, maar in hoe 'n mate Botha se redaksionele leiding die sentrale teks beïnvloed het, sal seker in 'n groter mate blyk met die verskyning van deel XIII (R). Volgens Botha (WAT XII: i) sou deel XII reeds in 2003 verskyn het, maar die totstandbrenging van die elektroniese weergawe van WAT I–XI sowel as die publikasie van *Etimologiewoordeboek van Afrikaans* en die elektroniese weergawe daarvan het deel XII se publikasie met twee jaar aangeskuif. Wat verder tot die vertraging aanleiding gegee het, is dat die WAT "gedwing is om drastiese besparingsmaatreëls in werking te stel en gevolglik is deel XII deur 'n kleiner redaksie as vorige dele saamgestel" (Botha in WAT XII: i). Die kleiner redaksie is die gevolg van 'n afgeskaalde staatsbydrae: die totstandkoming van elf staatsondersteunde woordeboekeenhede vir al die amptelike landstale in 2000 in plaas van net twee soos in die vorige taalbedeling het tot hierdie vermindering in die staatsubsidie gelei. Die Redaksie van die WAT verdien dus lof vir dit wat hy sedert 2003 bereik het, veral wat betref die tegnologiese vooruitgang met die verskyning van die elektroniese weergawe van die WAT.

In hierdie resensieartikel word die missie van die WAT as uitgangspunt geneem en word vasgestel in hoe 'n mate deel XII hieraan voldoen:

Die missie van die WAT is om die Afrikaanse taal en die gebruikers daarvan te dien, asook die leksikografie in die algemeen, veral die Suid-Afrikaanse leksikografie.

Die ideaal is dus dat die akademikus die WAT as die praktiese vergestaltung van die metaleksikografie kan benut, terwyl die deursneegebruiker die Afrikaanse taal so volledig moontlik daarin weerspieël wil sien: Standaardafrikaans, omgangstaal, streektaal, variëteite, taal vir spesifieke doeleindes (vak-

taal), uitdrukkings en kollokasies moet almal opgeneem en leksikografies bewerk word. Dit geld natuurlik nie net deel XII nie, maar daar wil veral nagegaan word of die veranderinge in deel XII wel leksikografiese en metaleksikografiese waarde toegevoeg het (Gouws 2006: 49).

Raamstruktuur

Binne die moderne metaleksikografie word woordeboeke beskou as "draers van tekssoorte" en die sogenaamde raamstruktuur word gebruik om die verskillende datatipes in 'n woordeboek te akkommodeer. Die woordeboekstruktuur kan in drie hoofdele verdeel word, naamlik die voortekste, die sentrale teks en die agtertekste. In die moderne leksikografie word *woordeboeke* in plaas van *woordelyste* nagestreef. Dit beteken dat die fokus nie meer net op die sentrale teks is nie, maar dat die buitetekste (voor- en agtertekste) saam met die sentrale teks gebruik word om woordeboeke na behore te ontsluit. Die buitetekste is veronderstel om die sentrale teks te komplementeer oftewel die raam te vorm waarbinne die sentrale teks geplaas en gelees word (Gouws en Prinsloo 2005: 57). Die buitetekste dra nie net by tot beter inligtingsontsluiting vir die gebruiker nie, maar speel ook 'n rol in die dataverspreidingstruktuur van die woordeboek (Keyser 2003: 35-36). Dit beteken dus dat deeglike beplanning nie meer net nodig is vir die strukturering en inhoud van die sentrale teks nie, maar ook vir die buitetekste.

Voortekste

Die voortekste en veral die teks wat die toeligting by die gebruik van die woordeboek bevat, word vandag as verpligte teks beskou. Geen leksikograaf kan aanneem dat die gebruiker van 'n spesifieke woordeboek sal weet hoe om die woordeboek te gebruik, te interpreteer en te verstaan, asook hoe om al die leksikografiese data na behore aan die sentrale teks te onttrek nie (Gouws en Prinsloo 2005: 57). Dit is des te meer belangrik by 'n omvattende woordeboek soos die WAT met sy uitgebreide data- en funksiebestek.

Die voortekste in die WAT beslaan 24 bladsye en bestaan uit 'n inhoudsopgawe, 'n inleiding, toeligting by die gebruik van die woordeboek, 'n skematiese voorstelling van die inskrywings en redaksionele afkortings. Die voortekste (uitgesonderd die inleiding) kan gesien word as geïntegreerde buitetekste, omdat hulle die gehalte van die inligtingoordrag van die WAT verhoog. Sien verder Gouws en Prinsloo 2005: 58-61 oor geïntegreerde en niegeïntegreerde buitetekste. Hier word ook 'n mediostrukturele verbintenis tussen veral die teks "Toeligting by die gebruik van die woordeboek" en die sentrale teks bewerkstellig deurdat die gebruiker met behulp van soms heel eksplisiete merkers na die sentrale teks gelei word. So byvoorbeeld kan die uitspraakslutel wat hier verskaf word, help om die fonetiese transkripsies in die sentrale teks te interpreteer.

Die "Toeligting by die gebruik van die woordeboek" (WAT XII: vii-xxi) wat die omvattendste voorteks is, bevat die afdeling "Aard van die woordeboek" waaruit die woordeboek se tipologie en die basis vir die opname van die lemmas blyk. Hierdie afdeling verduidelik ook die rangskikking van die verskillende betekenisonderskeidings binne een woordeboekartikel of tussen twee homonimiese lemmatekens. Daarop volg 'n indeks wat die volgende onderafdelings lys: 'n terminologieverklaring wat die leksikografiese vakterme verduidelik soos hulle in die res van die toeliggende aantekeninge gebruik word, die rangskikking van die woordeboek- en artikelinskrywings, samestellings en afleidings, die gebruik van hakies, verwysings, die beleid vir die hantering van beledigende en sensitiewe leksikale items, afkortings, bronne, uitspraaksleutel, en die missie van die WAT.

Wat taamlik hinderlik is, is dat die koherensie in die "Toeligting by die gebruik van die woordeboek" versteur word deur die invoeging van die "Skematiese voorstelling van inskrywings" op bll. viii en ix. Die skematiese voorstelling is uiteraard die een voorteks waarsonder geen woordeboek kan wees nie, maar hier moes dit eerder (soos in deel XI) voor of direk na die "Toeligting by die gebruik van die woordeboek" geplaas gewees het. Soos dit nou gehanteer is, is daar ook geen inskrywing in die indeks wat die gebruiker na die skematiese voorstelling lei nie. Daar is egter 'n hegte integrasie tussen sentrale teks en voorteks met die gebruik van inskrywings uit die sentrale teks as illustrasiemateriaal om die toeligting prakties te demonstreer. Daar word ook verder 'n hegte kohesie bewerkstellig binne hierdie voorteks deurdat daar by elke term wat in die skematiese voorstelling gebruik word, 'n verwysing na die betrokke paragraaf in die "Toeligting by die gebruik van die woordeboek" gegee word. Sonder hierdie leksikografiepraktyk sal die gebruiker nie werklik die woordeboekartikel na behore kan ontsluit nie. Soos in deel XI word die betrokke gedeeltes wat binne die woordeboekartikel benoem word, in die skematiese voorstelling ingekleur, wat die gebruiker help om die benoemde gedeelte makliker en vinniger te sien.

Wat die res van die "Toeligting by die gebruik van die woordeboek" betref, word daar feitlik deurgaans goeie integrasie tussen die sentrale teks en die voorteks bewerkstellig deurdat daar net soos by die skematiese voorstelling toepassings uit deel XII gebruik word. Daar is wel twee gevalle waar dit nie gebeur nie, naamlik onder 2.6 "Etiket" op bl. xiii: "lemmas wat kragwoorde is wat 'n spreker of skrywer se oordeel verraai" en "lemmas wat kragwoorde en uitroepe verteenwoordig wat gebruik word met die opset om te skok". By eersgenoemde word 'n voorbeeld uit Deel XI **naai II 2** en by laasgenoemde 'n voorbeeld uit Deel I **bliksem** verskaf. Dit is onduidelik waarom, want gepaste voorbeelde behoort maklik in deel XII self gevind te kan word. Onder parr. 5.4 "Verwysingsmetode by teenstellings en ander semanties verwante vorme", 5.5 "Verwysing met 'Vgl.' en die aanwending van 'OPM.'", en 5.6 "Verwysing met 'Sien Toeligting, 6'" word daar geen toepaslike illustrasies uit die sentrale teks gegee nie, terwyl daar konsekwent toepassings by die res van die toeliggende paragrawe gevind word. Heel aan die einde van die "Toeligting by die gebruik

van die woordeboek" ná "Verwysings" word daar 'n paragraaf aangetref met as opskrif "LET OP" (bl. xix). Dit word gelees as sou dit deel wees van die vorige par. 5.6, terwyl dit eintlik geen verband daarmee het nie. Dit handel wel oor verwysings, maar spesifiek oor waar verwysings na lemmas met syfers en punte vooraan, byvoorbeeld *35 mm-kamera* en *.303-geweer*, gesoek moet word. Hierdie inligting hoort logieserwys onder par. 2.3 "Die lemma en sy vorm" waar alle inligting oor die ordening van lemmatekens verskaf word, asook inligting oor verwysings na gemerkte lemmatekens: o.a. lemmatekens met syfers voor ter onderskeiding van homonieme en homograwe en meerwoordige leksikale items. Op sy huidige plek is dit feitlik onmoontlik vir die gebruiker om die inligting oor die plasing van hierdie soort lemmas met behulp van die indeks op te spoor.

Die gebruik van tipografiese en nietipografiese strukturele aanwysers as deel van die interne toegangstruktuur van die woordeboek word aangewend om die gebruiker na spesifieke tipes items binne die woordeboekartikel te lei. Volgens Gouws en Prinsloo (2005: 171) moet hierdie stelsel in besonderhede uitgespel word in die toeliggende gedeelte van die woordeboek. In die toeliggende van die WAT word net die gebruik van ronde en vierkantige hakies as afsonderlike indeksinskrywing volledig beskryf. Die gebruik van ander nietipografiese strukturele aanwysers soos die \approx en die \blacksquare word gehanteer as deel van die "Rangskikking van die woordeboek- en artikelinskrywings" onder die hofie "Voorbeelde" in par. 2.1 "Volgorde en tipografie" op bl. xi. Tipografiese en nietipografiese strukturele aanwysers word hier saam aangetref met die beskrywing van die gebruik van sowel kursiewe inligting as die aanwysers \approx en \blacksquare soos ook in par. 2.10 "Voorbeeldmateriaal". Die WAT behoort meer gebruik te maak van sowel tipografiese as nietipografiese strukturele aanwysers om die gebruiker vinnig en maklik na 'n spesifieke item of datakategorie binne 'n woordeboekartikel te lei sonder dat die hele artikel eers deurgelees moet word. Die enigste nietipografiese merker wat tans so gebruik word, is die vierkantige hakies wat uitsluitlik die uitspraak van 'n betrokke lemmateken aandui. Oorsien hierdie strukturele aanwysers in die WAT nie onderskeidend nie, aangesien dieselfde aanwyser aangewend word om heelwat datakategorieë te onderskei: vergelyk hier die verskillende gebruike van die ronde hakies wat aangetref word in par. 4.1 (bll. xvi-xvii). Net so word kursiefdruk nie uitsluitlik vir voorbeeldmateriaal gebruik nie en die aanwyser \blacksquare vir uiteenlopende tipes redaksionele kommentaar, weliswaar almal in aanloopvorm. Die strukturele aanwyser moet eerder die tipes kommentaar onderskei as die vorm waarin die kommentaar aangebied word. Die tipes kommentaar wissel vanaf die grammatikale soos die aanduiding van die meerkonksionele gebruik van 'n lemma (Ook selfst./ww./byw.) tot die aanwysing van 'n meer stilistiese toepassing van die lemma (In 'n ou bron/Met personifikasie). Grammatikale en stilistiese leiding behoort juis nie op dieselfde manier behandel te word nie. Styl is die keuses wat 'n taalgebruiker uitoefen, onderhewig aan of begrens deur die taalsisteem oftewel die grammatika. Die aanwyser \approx word gebruik om kollokasies en poëme van sitate te onderskei, maar wat onderskei kollokasies en poëme van

mekaar? Aangesien alle voorbeeldmateriaal gekursiveer word, kan nietipografiese strukturele aanwysers met vrug aangewend word vir die eenduidige onderskeid van kollokasies, poëme en sitate.

Die integrasie van die voortekste en die sentrale teks sal duideliker wees indien die bladsynommers van die voortekste nie afsonderlik van die sentrale teks genommer word nie. Die voortekste word met klein Romeinse syfers, en die sentrale teks met Arabiese syfers genommer.

Agtertekste

Die eerste agttertaks vanaf bl. 722 bestaan uit "Bronne waaruit in hierdie deel van die Woordeboek aangehaal is". As gevolg van die talle verwysings vanuit die artikels in die sentrale teks na hierdie agttertaks kan dit as 'n geïntegreerde agttertaks beskou word. Volgens Wiegand (1996: 13, soos aangehaal in Gouws 1999: 8) maak verwysings na die bronne waaruit aanhalings vir voorbeeldmateriaal geneem is, ook deel uit van die mediostruktuur van die woordeboek. Hierdie verwysings kan die gebruiker na 'n eksterne bron lei (Gouws 1999: 9). Die integrasie van die sentrale teks met die agttertaks blyk duidelik uit die deurlopende bladsynommering in Arabiese syfers vanaf die sentrale teks asook "Verbeterings in deel XI van die Woordeboek" wat vanaf bl. 760 aangetref word. Laasgenoemde gedeelte hoef eintlik nie by deel XII geïntegreer te wees nie, want dit hou geen verband met enigiets wat vir deel XII ter sake is nie. Deurdat dit wel as agttertaks voorkom en kennelik by deel XII geïntegreer is, word kohesie tussen die afsonderlike dele van die WAT bewerkstellig en sien die gebruiker dat die dele gesamentlik gebruik behoort te word.

Uit die tipologiese aard van die WAT, naamlik dié van omvattende verklarende woordeboek wat ten doel het om die Afrikaanse taal so volledig moontlik te weerspieël, word feitlik alles in die sentrale teks aangetref, en daarom voel die Redaksie waarskynlik dat daar nie behoefte is aan meer uitgebreide agttertaks nie. Volgens Gouws en Prinsloo (2005: 61) sou die beskrywing en gebruik van 'n reeks leksikale items uit byvoorbeeld 'n tegniese of ander gespesialiseerde veld wat by beperkter verklarende woordeboeke nie normaalweg as deel van die sentrale teks sou optree nie, as agttertaks geplaas kon word. So word daar in die *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) (2005) 'n lys afkortings sowel as geografiese name en hulle afleidings as agttertaks aangetref. Die gebruik van die agttertaks is hier dus aanvullend tot die sentrale teks. Niks verhoed die Redaksie van die WAT om byvoorbeeld afkortings sowel in die sentrale lys as in 'n agttertaks te plaas nie. Dit gee dan aan die gebruiker 'n keuse van toegangsroetes: óf die inligting word onttrek deur die alfabetiese sentrale lys te besoek, óf die gebruiker gaan direk na die agttertaks en kry die inligting gouer en makliker daar. Indien die gebruiker omvattender inligting sou wou hê oor die lemmateken in die agttertaks, word daar dan weer na die sentrale teks geblaai. Dit sou die waarde van verdere dele van die WAT leksikografies verhoog deurdat die woordeboek dan politoe-

ganklik vir sy gebruikers sal wees en daar dan ook besonder hegte integrasie tussen sentrale en agterteks bewerkstellig word. Nog gevalle van agtertekste sou byvoorbeeld etimologiese bewerkings van sommige lemmatekens asook sinoniem- en antoniemlyste kon wees.

Sentrale teks

Die sentrale teks wat gewoonlik as alfabetiese lys aangebied word, is die belangrikste deel van enige woordeboek, want dit is wat die algemene gebruiker as "die woordeboek" beskou. Die opmerkings in die volgende drie afdelings oor sommige artikels binne die artikeltrajek van deel XII vind plaas aan die hand van drie strukturele komponente wat in die sentrale lys bewerk moet word, naamlik die makrostruktuur, die mikrostruktuur en die mediostruktuur. Vir 'n bespreking van verdere strukturelemente en die terme *artikeltrajek* en *artikel*, sien Gouws en Prinsloo (2005: 62-64).

Makrostruktuur

Hierdie strukturele komponent het te make met die leksikale items wat uitgesoek word om as lemmatekens in die woordeboek opgeneem te word. Die makrostrukturele komponente word dan ook die belangrikste eenhede wat binne enige woordeboek leksikografies gehanteer word (Gouws en Prinsloo 2005: 63). Dit is die komponent waarvoor die WAT gewoonlik hoog deur resensente aangeprys word. Vergelyk byvoorbeeld:

Woordeskatte kan uitgebrei word met woorde soos *pinakoteek ...*, *pinas ...*, *proro-geer ...*, *psigogenese ...* en *puerpera ...* (Martins 2005)

Die letter P bevat 8 873 verklaarde lemmas, waaronder kostelike en beskrywende uitdrukkings soos 'paddakombers' ..., 'pampoemies' ..., 'pepa'..., 'petieterig', 'pikewynpak' ... en 'penswinkel' ... Die WAT is die enigste woordeboek wat nie net Standaardafrikaans reflekteer nie, maar die hele spektrum van die Afrikaanse woordeskat in sy geheel boekstaaf; van vaktaal met 'n hoë gebruiksfrekwensie, tot algemene gesels- en spreektaaluitdrukkings, variëteite van Afrikaans soos Griekwa-Afrikaans en Oranjerivierafrikaans, asook die Khoi- en Sanbydraes tot die Afrikaanse woordeskat. (Swart 2006)

Die omvattendheid van die WAT word verder deur Botha as Hoofredakteur in verskeie forums voorgehou. Een hiervan (Botha 2006) lui:

Die taak van die *Woordeboek van die Afrikaanse Taal* bly steeds om die ganse woordeskat van Afrikaans in al sy verskyningsvorme van standaardtaal, streektaal, geselstaal en die verskillende variëteite te probeer dokumenteer. Die WAT is die enigste woordeboek wat streektaal, geselstaal, uitdrukkings en die verskillende variëteite omvattend opneem. Die kleiner woordeboeke beperk hulle tot die standaardtaal.

Ongeag hoe omvattend die makrostruktuur daar ook al uitsien, die keuse van leksikale items wat hierdie omvattendheid van die Afrikaanse leksikon illustreer en verteenwoordig, mag nie willekeurig en lukraak gemaak word nie. "The lemmata have to be drawn from a representative corpus of the specific language and a vital part of the planning of any lexicographic endeavour is the development of a corpus," aldus Gouws en Prinsloo (2005: 65). Hulle sê voorts daar moet goed gedefinieerde beginsels wees op grond waarvan leksikale items as lemmas opgeneem word. Gouws (1999: 4) merk op dat daar in die beginstadias van die WAT nie juis blyke was van grondige leksikografiese beplanning nie. In die huidige "Toeligting by die gebruik van die woordeboek" blyk daar weliswaar steeds nie 'n plan vir opname nie, maar op bl. vii word dit gestel:

Afgesien van taal uit geskrewe tekste, word ook spreek- en streektaal gereflekteer indien daar genoegsame getuienis daarvan bestaan. Verder word terme wat in koerante en tydskrifte aangetref word, ook opgeneem. Die kriteria vir opname in al die gevalle is of 'n leksikale item gevestig geraak het in Afrikaans.

Die korpus waaruit die leksikale items geselekteer word, bestaan uit 'n kaartversameling wat sedert 1911 deurlopend opgebou word deur 'n groep medewerkers en wat sowel geskrewe as gesproke taal bevat, asook uit 'n elektroniese databasis wat saamgestel word met die oog op die alledaagse en frekwent gebruikte leksikale items.

Dat die bestaande korpus en/of die wyse van onttrekking en seleksie feilbaar is en herbedink behoort te word, blyk uit die weglating van items wat daagliks in gesproke en geskrewe taal voorkom. Die item **plaasmoord** kom nie as verklaarde of onverklaarde lemmateken voor nie. Van die vier items **plaasaanval**, **plaasarres**, **plaastronk** en **plaaswag** wat wel opgeneem is, is dit net **plaasaanval** en **plaaswag** wat naastenby so frekwent soos **plaasmoord** voorkom. Die items **plaasarres** en **plaastronk** se mindere frekwensie blyk daaruit dat hulle onderskeidelik as *histories* en *verouderend* geëtiketteer is, maar desondanks word hulle volle verklaarde lemmastatus gegee. Net so word **prettrap** met sy persoonsvorm **prettrapper** ook nie opgeneem nie, maar **prettdraf**, **prettdrawwer**, **pretloop** en **pretstap** wel. Op grond waarvan is die net so gebruiklike **prettrap** en **prettrapper** geïgnoreer? Twee hoogs frekwent gebruikte leksikale items wat ook teen alle verwagting in as verklaarde of onverklaarde lemmas ontbreek, is **padongeluk** en **padgeweld**. Die afkorting **POF** word wel gegee met as definieëns die volvorm "Padongelukkefonds", wat, omdat dit 'n eienaam is, nie as lemma opgeneem en verklaar is nie. Die rede vir die nieopname word egter nie in die "Toeligting" verduidelik nie.

De Schryver (2005: 70) merk ook die probleem van die weglating van frekwent gebruikte leksikale items op. Hy meen die gebruik van slegs 'n korpus is nie voldoende vir die onttrekking van leksikale items vir plasing in 'n omvattende woordeboek soos die WAT nie, selfs al is die hoofkriterium vir opname volgens Botha (2005: 86) optekening uit 'n minimum van drie bronne gebruik deur drie verskillende skrywers:

Hanks (2004) remarks ...: 'For a big dictionary — even a one-volume big dictionary, such as the (*New*) *Oxford Dictionary of English* — a corpus alone is not enough. ... We included *all* the words and senses in the corpus *and then some more*.' Given the *overall*-descriptive nature of WAT this means that ... (over 60% of the lemma signs found in HAT but missing in WAT) should at least have been given lemma-sign status in WAT.

De Schryver (2005: 72) beweer verder dat die nieopname van frekwent gebruikte leksikale items in die WAT daartoe kan lei dat die WAT verder en verder wegbeweeg van sy doel om 'n omvattende beskrywende woordeboek te wees en dat wat Botha steeds in deel XII voorhou, naamlik om "die ganse woordeskat van Afrikaans in al sy verskyningsvorme ... te probeer dokumenteer" deesdae nie meer gestand gedoen word nie. Vergelyk ook die opmerking van Feinauer (1996: 238-239):

Tot tyd en wyl die databasis van die WAT in alle opsigte die sosiolinguistiese werklikheid van die Afrikaanse taalgemeenskap reflekteer, sal die samestellers van komende dele hulle nie ten volle op die databasis kan verlaat, of hulle daardeur kan laat intimideer nie, en sal op eie inisiatief ook lemmas moet opneem. Die eerste prioriteit moet sekerlik wees om 'n meer omvattende woordeboek daar te stel, en nie om aan die databasis getrou te wees nie.

Volgens De Schryver (2005: 72) is 'n moontlike oplossing vir hierdie probleem die volgende:

A large electronic corpus should be built and *effectively used* to complement the current manually compiled citation collection. This will avoid a situation whereby further frequent items are not given lemma-sign status.

Op hierdie stadium lyk dit asof die kaartversameling nog te sterk bevoordeel word bo die elektroniese korpora wat die Redaksie tot sy beskikking het. Dit lei daartoe dat te veel aandag aan niefrekwente leksikale items gegee word ten koste van hoogs frekwente items. Daar sal later na hierdie punt teruggekeer word in die bespreking van die mikrostruktuur.

Nog 'n probleem wat De Schryver (2005: 66-68) uitlig, is die lyste en lyste onbewerkte lemmatekens oftewel sublemmatekens wat in deel XI aangetref word. Volgens hom lyk dit asof die WAT deesdae met die woordelysgedeelte van die *Afrikaanse Woordelys en Spelreëls* (2002) vergelyk kan word. Dit behoort volgens hom nie te gebeur in 'n tydperk van elektroniese korpora dat hoogs frekwent gebruikte items bloot as sublemmas in lang lyste verskyn sonder enige bewerking nie. Van die lemmas is volkome deursigtig en kan dus met min probleme onverklaard opgeneem word met slegs klem- en woordsoort-aanduiding. Vergelyk hier **pop** met die volgende onverklaarde lemmas: **popbed**, **popbottel**, **popfabriek**, **popfiguur**, **pophare**, **popklere**, **popkometers**, **popmuseum**, **poprok**, **popsteestel**, **popwaentjie** en **popwiegie**. Leksikografe kan egter nie altyd aanvaar dat wat vir hulle deursigtig is, ook vir alle gebruikers ewe deursigtig sal wees nie. Die volgende sublemmas word as voorbeelde genoem van talle ander wat eerder in deel XII verklaar moes gewees het omdat

hulle nie sonder meer deursigtig is nie: **padaflewing** (Kan alle gebruikers aflei dat dit aflewing is wat per vragmotor geskied?), **plaaskombuis** en **plaaskultuur** (Hierdie lemmas het veral die afgelope tyd polisemiese uitbreiding ondergaan na onderskeidelik 'n sekere tipe restaurant en 'n sekere tipe kultuur), **polisiewag** en **polisiewreedheid** (Is eersgenoemde 'n wag of dalk 'n erewag deur die polisie of selfs 'n wag vir die polisie? Is laasgenoemde wreedheid deur of teenoor die polisie?), **poskoets** (Dit is 'n historiese item wat min vir die hedendaagse gebruiker beteken en verklaring verdien.).

Die een woordeboek waarin daar verklaar kan en moet word, is die omvattende beskrywende woordeboek. Omvattende woordeboeke is juis veronderstel om tweeledig omvattend te wees: sowel makro- as mikrostruktureel. Op makrostrukturele vlak word daar dan wel voldoen aan die omvattendheidsbeginsel met die opname van onbewerkte lemmas, maar op mikrostrukturele vlak skiet die WAT dikwels tekort. Die enkele en enigste rede waarom lemmas onbewerk gelaat behoort te word, is op grond van hulle semantiese deursigtigheid al dan nie. Betekenisverklaring is maar slegs een van die artikel-inskrywings waarvoor die gebruiker die woordeboek raadpleeg. Indien 'n bepaalde lemmateken onbewerk gelaat sou word selfs al sou die betekenis daarvan deursigtig wees, word die gebruiker steeds ontnem van die res van die mikrostrukturele bewerking wat al die linguistiese kenmerke van die betrokke lemmateken weerspieël. Vergelyk hier ook Gouws en Prinsloo (2005: 49) se kommentaar oor die mikrostruktuur van die omvattende woordeboek (Beklemtoning deur die skrywer):

The microstructure gives an extensive account of the linguistic features of the lemma signs. This leads to a *balanced* data density, i.e. the quantitative relation between macro- and microstructural entries, in the dictionary.

Die WAT sal in komende dele moet waak teen 'n ongebalanseerde oorwig aan die kant van die makrostruktuur ten koste van die mikrostruktuur. Daar word later na die probleem van onbewerkte lemmas teruggekeer in die bespreking van die mediostruktuur.

Mikrostruktuur

Die mikrostruktuur verwys na die onderskeie datakategorieë wat in die hantering van die lemmateken voorkom. Die makrostrukturele element tesame met die mikrostrukturele hantering daarvan maak 'n woordeboekartikel uit. Elke afsonderlike item waaruit die woordeboekartikel saamgestel is, staan bekend as 'n inskrywing (Gouws en Prinsloo 2005: 64 en 115). Van den Berg (2006) verwys dus na die mikrostruktuur van die WAT wanneer hy skryf:

Die *Woordeboek van die Afrikaanse Taal* (WAT) is meer as net 'n woordeboek. Dis daarby 'n lewenshandleiding, want dit verklaar nie net woorde abstrak nie, maar demonstreer konkreet wat mense hoe, waar en wanneer oor die lewe gesê het.

Die "hoe" kan byvoorbeeld verwys na die inskrywings uitspraakaanduiding of die sitate en poëme onderskeidelik as konteks- en koteksleiding vir die gebruiker. Kontekstuele inskrywings help die gebruiker in sowel teksresepsie as teksproduksie, terwyl kotekstuele inskrywings die gebruiker sterker in teksproduksie bystaan deur middel van kollokasies, frases, en voorbeeldsinne wat aantoon hoe die lemma binne die spesifieke taalsisteem optree (Gouws en Prinsloo 2005: 128-129). Die "waar" en die "wanneer" kan weer verwys na die etikette wat die geografiese gebied of die gebruiklikheid of tydgebondenheid van 'n lemma aandui, wat albei eksplisiet kontekstuele leiding gee, maar wat ook uit die sitate afgelei sou kan word. Soos Van den Berg dus tereg opmerk, verskaf die WAT meestal pragmatiese leiding oftewel "konkrete demonstrasie" aan die woordeboekgebruiker deur middel van 'n uitgebreide en omvattende reeks inskrywings wat in die "Toeligting by die gebruik van die woordeboek" gedemonstreer en verduidelik word.

Die mikrostruktuur van enige woordeboek hang ten nouste saam met die hoof-leksikografiese funksie van die woordeboek. Volgens Bergenholtz en Tarp (2003: 176) verteenwoordig leksikografiese funksies die hulp wat die woordeboek aan die gebruikers bied. Hierdie leksikografiese funksies is die basiese elemente waarop die teorie en praktyk van die leksikografie berus en hulle is die hoofbeginsel van die inhoud en vorm van woordeboeke. Die hele woordeboekstruktuur, maar by uitstek die mikrostruktuur, word aan die hand van die betrokke funksies gerig. Gesamentlik vorm al hierdie funksies en die veld wat die woordeboek dek, die werklike doel van die betrokke woordeboek. Bergenholtz en Tarp (2003: 172) maak 'n indeling tussen twee hoofgroepe leksikografiese funksies wat hoofsaaklik gegrond is op die behoefte van die woordeboekgebruiker, naamlik kognitiewe of kennisgerigte funksies en kommunikatiewe funksies. Die kognitiewe funksies verskaf onder andere algemene kulturele, ensiklopediese en etimologiese inligting, meer spesifieke inligting oor die register of vakgebied van die lemmateken, asook inligting oor hoe die lemmateken binne die taalsisteem optree (o.a. grammatikale leiding). Kommunikatiewe funksies wat hier ter sake is, is dié wat verband hou met teksproduksie en teksresepsie (in een of meer tale). Albei funksies is wel ter sake vir 'n omvattende woordeboek soos die WAT, maar miskien is die hoof funksie hier tog meer kognitief as kommunikatief. Vir laasgenoemde funksie sal die gebruiker eerder standaard verklarende woordeboeke, aanleerwoordeboeke of vertalende woordeboeke raadpleeg. Volgens Gouws (1999: 5-6) is die tipiese gebruiker van 'n omvattende woordeboek 'n moedertaalspreker met 'n bogemiddelde beheersing van die taal; die potensiële gebruikers van die WAT is dus meestal bogemiddelde, oor die algemeen meer gesofistikeerde Afrikaanssprekende woordeboekgebruikers.

Gouws en Prinsloo (2005: 14-15) beklemtoon dat die vasstelling van die leksikografiese funksie en voortvloeiend daaruit die hoofdoelstelling van die woordeboek 'n sentrale rol behoort te speel in die konseptualisering en beplanning van die woordeboek. WAT XII is skynbaar nog nie aan die hand van lek-

sikografiese funksies gekonseptualiseer en beplan nie. Volgens De Schryver (2005: 62-63) lyk dit asof die Redaksie korter en minder ensiklopediese definiense skryf asook asof etimologiese inligting verminder word, met ander woorde daar word aan die kognitiewe inligting besnoei. Aan die ander kant is daar 'n uitbreiding van uitspraakleiding, een van die tipiese kommunikatiewe funksies wat vir mondelinge teksproduksie nuttig is. Vergelyk hier Gouws en Prinsloo (2005: 120) se opmerking (Beklemtoneering deur die skrywer):

Where text reception or text production are the *dominant functions* and the dictionary has to assist users in oral communication, a thorough treatment of the sound aspects of a given word is needed.

Vir die WAT as meer kognitief gerigte woordeboek moes die verskuiwing dus net andersom gewees het: meer etimologiese inligting en meer ensiklopediese definiense en 'n inkorting van uitspraakleiding. Die uitspraakleiding van die lemmateken **protea** beslaan byvoorbeeld feitlik twee reëls: [prote:^ha / prute:^ha / prœute:^ha / proti:^ha / pruti:^ha / prœuti:^ha] en die volle reeks uitspraakmoontlikhede word by alle verdere lemmatekens met **protea** herhaal: **Protea**, **proteablom**, **proteaboom**, **proteabos**, **proteafamilie**, **Proteakleure**, **proteaplant**, **Proteas**, **proteasoort**, **proteaspesie**, **proteastruik**. Hierdie ruimte wat deur die uitspraakleiding opgeneem word, sou met veel groter vrug kognitief aangewend kon word.

As die WAT dan wel implisiet by die saamstel van deel XII besluit het om die woordeboek meer kommunikatief te rig, moes die uitspraakleiding dan nie bloot deskriptief gelys gewees het nie, maar moes daar 'n mate van normering gewees het om te voorkom dat die gebruiker onwetend 'n minder "korrekte" uitspraak sal gebruik. By die lemmateken **prestige** word die volgende uitsprake gelys: [præsti:ʒ' / præsti:ʒ' / præsti:dʒ' / præsti:dʒ' / præsti:tʃ' / præsti:tʃ']. Moet die gebruiker nou aflei die gebruiklikste uitspraak word eerste en die mins gebruiklike laaste gegee? Ook hier word inkonsekwent met uitspraakaan duiding omgegaan, want uit die hantering van **protea** en **prestige** sou die gebruiker verwag om deurgaans alle uitsprake aan te tref. By **polonie** byvoorbeeld word die Engelse uitspraak [pələu'ni] wel gegee, so ook by **pro-am** [prœu'æm]. By **polonek**, **polonekhemp** en **polonektrui** kom die heel gebruiklike Engelse uitspraak [pœu'lœu] egter nie voor nie. Dan kon daar ook geen patroon afgelei word by die uitspraakleiding by afkortings nie en dit word ook nie in die "Toeligting" verduidelik nie. Uit die volgende afkortings **pm**, **¹p.m.**, **²p.m.**, **³p.m.**, **Pm**, **PM**, **Pmb.**, **¹PMD**, **²PMD**, **PMG**, **PMM**, **PMS** word uitspraakleiding net by **PM**, **PMG** en **PMS** gegee. Dat sommige afkortings selde uitgespreek word, terwyl ander, veral akronieme, meestal uitspraakleiding moet ontvang, blyk nie uit hierdie behandeling nie. By ten minste **¹PMD**, **²PMD** en **PMM** moes ook uitspraakaan duidings gewees het.

De Schryver (2005: 62-63) merk op dat daar deesdae "an extreme disparity in average article length" in die WAT voorkom. Wat egter meer hinder aan die lengte van die artikels in deel XII, is dat minder frekwent gebruikte lemmas veel langer artikels as sommige hoogs frekwente lemmas beslaan. Dit kan weer

eens toegeskryf word aan gebrekkige beplanning by korpusonttrekking: frekwent gebruikte spreektaalvorme (veral dié met 'n hoër taboewaarde) word afgeskeep teenoor ongebruiklike skryftaalvorme. Vergelyk hier byvoorbeeld die lengte van die woordeboekartikels vir **poes** en **piel** met dié vir **predikaat**. By geeneen van eersgenoemde twee lemmatekens word daar etimologie aangedui nie. Daar word ook geen uitdrukkings by hulle gegee nie, terwyl daar wel sulke uitdrukkings in die spreektaal voorkom. By **poes** se eerste polisemiese inskrywing word net een sitaat van André P. Brink (1976) verskaf en dit in die diminutief *poesie*. By die tweede polisemiese inskrywing kom daar geen koteksleiding voor nie. By **piel** word daar as koteksinskrywings slegs twee sitate verskaf van Deon Meyer (1996) en weer eens van André P. Brink (1973) in die meervoudsdiminutief *pieletjies*. Die woordeboekartikel van **piel** beslaan 7,5 en dié van **poes** 8 reëls. Daarteenoor beslaan die lemmateken **predikaat** wat deesdae hoofsaaklik as linguïstiese term gebruik word een hele kolom. Dat die spreektaal kennelik afgeskeep word ten koste van die skryftaal blyk ook uit byvoorbeeld die omvang van die woordeboekartikels vir **poegie** en **poëem**. Vroeër reeds het Feinauer (1996: 258) opgemerk:

Hierdie probleem spruit uit die Prerenaisance-metode van ekserpering deur medewerkers, wat steeds in die WAT gevolg word: vreemde woorde word juis opgeteken om te bewys dat hulle bestaan, en heelwat van hierdie optekeninge word as voorbeeldmateriaal geplaas. Lombard (1992: 154) wys daarop dat ekserpeerders eerder die ongewone as die gebruikelike opteken. ... Die redaksie van die WAT moet bedag wees hierop en nie in verdere dele voortgaan met hierdie misleidende beeld t.o.v. die frekwensie van ongewone lemmas nie, of die WAT moet dit in sy toeligting baie duidelik stel dat sy sitate nie as absoluut verteenwoordigend van die taal beskou word nie, veral nie wat die frekwensie van woordgebruik betref nie.

Dat die vasstelling van gebruiklikheid 'n probleem is, blyk ook uit die opname van die lemma **poskaffer**, terwyl die sinchronies veel gebruikliker **plaaskaffer** nie opgeneem is nie. Onder "Beleid vir die hantering van beledigende en sensitiewe leksikale items" (bl. xix) verskyn die volgende: "Alle beledigende of sensitiewe simplekse, komposita en uitdrukkings wat aan die *algemene gebruiklikheidskriterium* van die WAT voldoen, word gelemmatiseer" (Beklemtoning deur die skrywer). Die WAT sal in die vervolg veel duideliker moet uitspel waarop hierdie algemene gebruiklikheidskriterium berus, want **poskaffer** moes nooit opgeneem gewees het nie; selfs met die etikette *rassisties* en *verouderd* (let wel) en die inskrywing "Sien Toeligting 6" sal dit gebruikers krenk. Onmiddellik ontstaan die vraag dan waarom die inskrywing "Sien Toeligting 6" nie ook by die lemmateken **poefter** voorkom nie? Hier verskyn wel die etikette *geselstaal*, *seksisties* en *neerhalend*. In die WAT-beleid word op bl. xix vermeld dat hierdie verwysing by rassismes aangedui word. Dit lyk dus asof die Redaksie rassisme as 'n growwer belediging en as meer onsensitief sien as seksisme, terwyl die Suid-Afrikaanse grondwet die twee gelykstel. Die beleid sê verder (bl. xx): "Waar moontlik, word 'n kwetsende leksikale item na die gebruiklikste neu-

trale sinoniem verwys. Indien dit nie moontlik is nie, word die definisie so neutraal as moontlik gestel." Reëlreg hierteenoor word daar by **poefter** 'n derde poliseem verstrekk met 'n verwysing "Moffie (³MOFFIE 2)" wat heelwat meer vernederend en kwetsend as *poefter* ervaar word.

In die artikels **paaltjieswagter** en **paaltjiewagter** word daar goeie kohesie tussen konteks- en koteksinligting bewerkstellig. By **paaltjieswagter** verskyn die etiket *verouderend* en die gebruiker word na die meer gebruiklike **paaltjiewagter** vir die definisie verwys. Die sitate wat by **paaltjieswagter** gebruik word, toon dat dit weliswaar verouderend is met een sitaat sonder jaar en nog een uit 1964. Dit wat die etiket aantoon, word deur die sitaat bevestig. By **paaltjiewagter** word ook 'n sitaat sonder jaar opgeneem, maar dan word 'n sitaat uit 1990 ook verskaf. Die WAT lei sy gebruikers egter nie kontekstueel deur die etikette *modern* of *neologisme* nie en die leser moet hierdie inligting meestal implisiet vasstel. So moet daar uit die koteks van **paaltjiewagter-kolwer**, naamlik die sitaatdatums 1998 en 1999, afgelei word dat dit 'n moderne woord is, asook uit die etiket *verouderend* wat by die sinoniem **kolwer-paaltjiewagter** gegee word. Etiket word egter nie konsekwent by verouderende lemmatekens gebruik nie. Net soos die gebruiker uit die sitaatdatums, 1995 en 2001, by **pixel** moet aflei dat dit 'n moderne toevoeging tot die Afrikaanse leksikon is, moet daar uit die sitaatdatums, 1951 en 1947, by **pitso** afgelei word dat dit 'n verouderende of weinig gebruiklike item binne die Afrikaanse leksikon is. Dit sou heelwat vinniger en dus gebruiklikervriendeliker kon geskied, indien hierdie inligting reeds as konteksleiding gegee sou word.

Op bl. vii verklaar die Hoofredakteur: "Die WAT is 'n omvattende woordeboek wat *oorwegend sinchronies* van aard is" (Beklemtoneering deur die skrywer). Die huidige korpus veroorsaak egter dat dit nie altyd so uit die sitate blyk nie. Die sitate weerspieël steeds nie die werklike gebruiksfrekwensie van lemmas nie, soos afgelei kan word uit die jaartalverspreidings wat nie verteenwoordigend van die betrokke lemmas se werklike gebruik is nie. Dit kom ook steeds voor asof sitate lukraak opgeneem en gerangskik word. Sowel Gouws (1994) as Feinauer (1996) het reeds vroeër daarop gewys. Vir die eerste twee polisemiese onderskeidings van die lemmateken **poësie** lyk die sitaatdatums soos volg: sonder jaar, 1959, 1932 en 1966 vir onderskeiding 1 waar die betekenis "Skriftelike taalgebruik in versvorm ..." verklaar word; 1966 en 1966 vir onderskeiding 2. Indien die WAT dan wel *sinchronies* gelees sou word, lei die gebruiker af dat die woord *poësie* in die tweede betekenisonderskeiding, "Poësie as literêre genre", in 1966 heel frekwent was, maar dat dit sedertdien in onbruik geraak het. Met die gebruik van die 1966-sitaat by poliseem 1 word die verouderde spelling van "nòg ... nòg" boonop onnodig geperpetueer. Eers vanaf die derde polisemiese inskrywing, die betekenis "Poëtiese werke van 'n bepoeter, taalgroep of nasie ..." word moderner sitate aangetref, naamlik 1932, 1993, 1974, 1983 en 1981. By die vierde polisemiese inskrywing waar die metaforiese gebruik "Iets wat 'n digterlike bekoring het ..." behandel word, lyk dit weer asof dit 'n verouderende gebruik is met sitaatdatums 1968, 1978 en 1935.

By polisemiese onderskeiding 5 "Studierigting, vak ..." word geen sitate aangetref nie, slegs poëme. Dit is 'n bewys dat die huidige korpus steeds nie voorsiening maak vir die algemene gebruik van woorde nie.

Mediostruktuur

Die mediostruktuur is die stelsel van alle tipes verwysing of kruisverwysing wat in 'n woordeboek aangetref word. Dit dui ook op die mate van tekstuele kohesie en koherensie wat daar tussen die onderskeie strukturele komponente binne 'n woordeboek voorkom (Gouws en Prinsloo 2005: 177 e.v.). Daar is vroeër reeds daarop gewys dat daar 'n mediostrukturele verhouding tussen die WAT se voortekste en die sentrale teks, sowel as tussen die sentrale teks en die agtertekste bestaan. Hier gaan hoofsaaklik gekonsentreer word op die mediostruktuur van die sentrale teks, maar daar gaan terugverwys word na die afdeling "Verwysings" (bll. xvii-xix) in die "Toeligting by die gebruik van die woordeboek" wat uitvoerig as voorteks beskryf word. Hier word 'n uiteensetting van al die onderskeie tipes verwysings in WAT XII gevind. In Gouws (1999) word 'n diepteondersoek gedoen na die mediostrukturele verteenwoordiging in WAT X en die meeste daarvan, positiewe sowel as negatiewe kritiek, geld mutatis mutandis vir WAT XII. Daarom gaan hier volstaan word met enkele opmerkings.

Wat sinoniemverwysing betref, m.a.w. waar die gebruiker die definiens by die frekwens gebruikte lid van die sinoniemparadigma moet gaan soek, word daar steeds geen eksplisiete tipografiese merker na die verwysingsadres gegee nie. Die gebruiker kan dit dalk verkeerdelik as 'n baie kort definiens in plaas van 'n sinoniemverwysing beskou, veral wanneer die verwysingsadres in 'n ander deel van die WAT aangetref word. Dit kan veral gebeur wanneer die gebruiker nie telkens verwys word na die voorteks om die onderskeie tipes sinoniemaanduidings te onderskei nie. Vergelyk **poepholvrug** en **poepketel** wat direk ná mekaar verskyn. By eersgenoemde word "Hoendereier" as sinoniemverwysing of definiens verskaf met 'n poëem wat tot 'n mate betekenisverklarend is: *Ek gaan soos 'n hen begin kekkel as ek nog poepholvrugte moet eet!*. By **poepketel** word "Ketel (¹KETEL 9)" as sinoniemverwysing of definiens aangetref met 'n poëem wat niks bydra tot die betekenisontsluiting van die lemmateken of die verwysingsadres nie: *Hy sit al weer op die poepketel met 'n boek*. 'n Gesofistikeerde woordeboekgebruiker sal hier wel uit die homonimiese en polisemiese aanduidings wat konsekwent in deel XII verskaf word, aflei dat dit 'n sinoniemverwysing moet wees, maar nie by **poepholvrug** nie. Volgens Gouws (1999: 15) "the WAT X system does not always enable the user to achieve a clear and quick identification of the reference relation. This is due to excessive textual condensation". Daarteenoor word daar wel 'n eksplisiete kruisverwysingsmerker "sin." ná gevalle van sinoniemvermelding gegee waar die verwysing nie so belangrik is vir betekenisontsluiting nie, want die definiens kom in die betrokke artikel voor:

prinsdom ... Gebied ... wat deur 'n prins ... regeer word; sin. *prinsipaliteit* ...

Waar WAT XII wel sy gebruikers behulpsaam is, is waar daar blinde verwysings na vorige dele sal wees. By **queen 3 b i** word byvoorbeeld na *fopdosser* verwys, maar *fopdosser* is nie in deel II as lemma opgeneem nie. Daarom word *fopdosser* saam met 'n definiens as sinoniemverwysing vir **queen 3 b i** aangebied. By **queen 3 b ii** word die gebruiker na die sinoniem *transvestiet* verwys, wat eers in 'n toekomstige deel van die WAT verklaar sal word. Weer eens kon die WAT die gebruiker hier heelwat meer behulpsaam gewees het, indien daar 'n meer betekenisverklarende poëem of sitaat gegee sou gewees het in plaas van die bestaande: 'n *Transvestiet in 'n vroulike blombloes en sandale was sigbaar oorstelp: "Kyk, hier's baie queens in die Kaap, maar dis die enigste royal ene"* (Beeld, 21 Mrt. 1995, 1). Nêrens kan hieruit afgelei word dat dit om 'n man in vrouekleure gaan nie. Die voorstel van Gouws (1999: 32) dat die Redaksie dit kan oorweeg om 'n strukturele merker by sodanige verwysings in te voeg om vir die gebruiker aan te toon dat die verwysingsadres nog nie bestaan nie, is nie opgevolg nie.

Gouws (1999: 17) maak beswaar daarteen dat WAT X nie makrostruktureel voorsiening gemaak het vir die meer frekwent gebruikte losgeskrewe *mooi praatjies* nie. Slegs die minder gebruiklike **mooipraatjies** word opgeneem, weliswaar met 'n kruisverwysing na **mooi praatjies**.

mooipraatjies ... Ook soms *mooi praatjies*. ...

Die losgeskrewe *mooi praatjies* is egter wel as idiomatiese uitdrukking onder die lemmateken **mooi** vermeld. Omdat daar nie 'n eksplisiete verwysing na *mooi praatjies* as idiomatiese uitdrukking verskyn nie, kan dié misverstand maklik ontstaan. Uit die toeligting onder "5.3 Verwysingsmetode by los- en vasgeskrewe vorme" (bl. xix) blyk dit dat dié beleid in WAT XII voortgesit word: "Losgeskrewe leksikale items (met uitsondering van uitdrukkinge) wat 'n vasgeskrewe teenhanger het, word *slegs by die vaste vorm* vermeld, en 'n aanduiding of die los vorm minder gebruiklik, gebruikliker of gelykwaardig is aan die vaste vorm, word gegee, bv. onder **pakkry**: Meer dikw. *pak kry*. Die definisie volg daarna" (Beklemtoning deur die skrywer). In die sentrale teks word **pakkry** egter soos volg gehanteer:

pakkry ... (*minder gebruiklik*)

1 Sien PAK KRY (³PAK, uitdr.) ...

2 Meer dikw. *pak kry*. Verloor of 'n nederlaag ly ...

Vir albei hierdie inskrywings word daar onder ³**pak** se gespesialiseerde uitdrukkinge vir polisemiese onderskeiding 1 en onder idiomatiese uitdrukkinge vir polisemiese onderskeiding 2 kruisverwysings aangetref met die definiens wat wel ook vir onderskeiding 2 by die gebruikliker losgeskrewe vorm herhaal word:

kry: pak kry Ook soms *pakkry*. 'n Pak (³PAK 1) toegedien word ...

kry: pak kry of soms **pakkry** Verloor of 'n nederlaag ly ...

Hier kan nie dieselfde misverstand as by **mooipraatjies** ontstaan nie, omdat die definiens herhaal word. Die misverstand ontstaan omdat daar nie eksplisiete verwysings by gespesialiseerde en idiomatiese uitdrukkings na die los- of vasgeskrewe vorme gegee word wanneer die definiense nie herhaal word nie, maar omdat hulle op dieselfde manier bewerk word as los- en vasgeskrewe leksikale items wat nie gespesialiseerde en idiomatiese uitdrukkings is nie. Die volgende geval kan as voorbeeld dien:

pypstop ...

II ww. ... Ook *pyp stop*.

Aangesien dit nie 'n eksplisiete kruisverwysing is nie, word *pyp stop* in die losgeskrewe vorm dus nie as lemmateken aangetref nie. Die WAT behoort 'n onderskeid te maak in die behandeling van dié twee soorte los- en vasgeskrewe items.

Om terug te keer na De Schryver (2005: 66-68) se beswaar teen die lyste en lyste onbewerkte lemmatekens wat in die WAT aangetref word: Van hierdie lemmatekens veroorsaak "semi-dead cross references" wanneer die gebruiker na 'n lemmateken verwys word, maar die lemmateken slegs as sublemmateken opgeneem is. Voorbeelde wat hy noem, is **kritiekloos** en **krities** waar die gebruiker in albei gevalle na die antoniemvorm **onkrities** verwys word, maar **onkrities** is dan nie verder bewerk nie en word slegs as sublemmateken onder die lang lys afleidings onder die prefiks **on-** aangetref. Dit is ook die geval by **krygskreet** waar die sinoniemverwysing **oorlogskreet** gegee word as die meer gebruiklike vorm, maar **oorlogskreet** word dan net as sublemmateken onder die lemma **oorlogs-** aangetref. Dieselfde tipe kruisverwysings kom in WAT XII voor met die onbewerkte lemmatekens wat onder **pre-** en **post-** gegee word. Die woordeboekartikels is baie informasiedig en gekondenseer as gevolg van die eksplisiete en implisiete kruisverwysende verbindings wat hier aangetref word:

post- 2 (*t.o.v. posisie*) Agter of posterior (POSTERIOR I 1) tot die geïmpliseerde voorwerp of saak; teenoor *pre-*, bv. *postaksiaal*, *postnasaal*, *postorbitaal*, *postsinapties*, *postverbaal*, *postwerkwoordelik*; **postalveolêr**, **postdorsaal**, **postduktaal**, **postganglionies**, **postkapillêr**, **postkraniaal**, **postnominaal**, **postpituïtêr**, **postvelêr**, **postvokalies**.

Uit die lys onbewerkte lemmas het slegs **postpituïtêr** 'n woordeboekinterne kruisverwysingsadres; die res van die verwysings is almal woordeboekekstern en **postvelêr** en **postvokalies** is boonop steeds dooie verwysings. 'n Gesofistikeerde woordeboekgebruiker behoort minstens die betekenis van die onbewerkte lemma **postpituïtêr** redelik maklik te kan aflei: "onverklaarde afleidings wat op grond van *deursigtigheid* nie verklaar hoef te word nie" (bl. xvi) (Beklemtoning deur die skrywer). Die mediostruktuur van die artikel sal ook help met die *deursigtigheid*: eerstens is daar die konteksleidende etiket (*t.o.v. posisie*); dan die woordeboekinterne kruisverwysing na (POSTERIOR I 1) as deel van die definiens, sowel as 'n antoniemverwysing (teenoor *pre-*). Daarop volg voorbeelde

wat ook as kruisverwysings kan dien, want hulle word verklaar (*postaksiaal, postnasaal, postorbitaal, postsinapties, postverbaal, postwerkwoordelik*), so deeglik dat **postverbaal** en **postwerkwoordelik** wat sinonieme is, albei as verklaarde lemmatekens opgeneem word. Hieruit kan dan afgelei word hoe die betekenis binne ander afleidings saamgestel is: **postnominaal** sal in dieselfde verhouding optree as **postverbaal** sowel as **postvelêr** teenoor **postnasaal**. Al wat nou oorbly, is om te kyk wat **pituïtêr** beteken, maar die betekenis is nog steeds nie duidelik nie. Is dit: 1. geleë aan die agterkant van die pituïtêre klier, 2. nadat die pituïtêre klier afgeskei het, of 3. nie een van die vorige twee nie. Die antoniem **pre-** lewer ook niks verhelderends nie, want **prepituïtêr** is nie opgeneem nie en bestaan waarskynlik glad nie indien betekenis 1 die korrekte sou wees. Alhoewel die mediostruktuur leksikografies deeglik uitgebou is, is dit steeds nie voldoende om die betekenis en nog minder die gebruik van alle onverklaarde lemmas te ontsluit nie. Vir toekomstige dele sal die Redaksie sy beleid van lang lyste sublemmas moet hersien en nie langer lukraak besluit wat verklaar moet word en wat nie soos bv. met **postorbitaal** (verklaard) en **postpituïtêr** (onverklaard) nie.

Slot

Niemand sal die leksikografiese verbetering van die WAT sedert die eerste agt dele ontken nie. Daarvoor verdien ook die Redaksie van deel XII lof. Soos enige wetenskap ontwikkel die metaleksikografie egter voortdurend. Die WAT sal in die toekoms nader aan die uitleef van sy missie kom, indien toekomstige dele deeglik gekonseptualiseer en beplan sou word aan die hand van leksikografiese funksies, veral noudat die metaleksikografie ook wegbeweeg vanaf die meer linguïstiese na die meer funksionalistiese benadering. Die keuse van leksikale items wat hierdie omvattendheid van die Afrikaanse leksikon illustreer en verteenwoordig mag ook nie langer willekeurig en lukraak gemaak word nie, maar 'n korpus moet beplan en gebruik word. Sodoende sal die Afrikaanse taal en die gebruikers daarvan beter gedien word. Frekwent gebruikte woorde sal dus nie langer as lemmatekens oor die hoof gesien word nie en die lengte van artikelinskrywings behoort ook meer gelykvormig te word. Die Suid-Afrikaanse leksikografie sal beter gedien word indien die WAT terugkeer na wat van 'n omvattende woordeboek verwag word, naamlik 'n meer gebalanseerde verteenwoordiging van makro- en mikrostruktuur. Dit behoort ook te lei tot 'n meer eweredige verspreiding van artikelomvang en minder onbewerkte lemmatekens.

Verwysings

- Bergenholtz, Henning en Sven Tarp. 2003. Two Opposing Theories: On HE Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.

- Botha, Willem.** 2005. Concurrent Over- and Under-treatment. A Short Reply. *International Journal of Lexicography* 18(1): 77-88.
- Botha, Willem.** 2006. Kom ons doen dit vir Afrikaans! Die WAT word 80. *Eikestadnuus*, 24 Maart.
- De Schryver, Gilles-Maurice.** 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): 47-76.
- Feinauer, Ilse.** 1996. Die negende deel van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 6: 233-271.
- Gouws, Rufus H.** 1994. *Woordeboek van die Afrikaanse Taal*, negende deel. SAUK, Skrywers en Boeke.
- Gouws, Rufus H.** 1999. Mediostructural Representation, Textual Condensation and User-orientation in the WAT X. *Lexicographica: International Annual for Lexicography* 15: 4-37.
- Gouws, Rufus H.** 2006. Die zweitsprachige Lexikographie Afrikaans-Deutsch — Eine metalexikographische Herausforderung. Dimova, Ana, Vida Jesenšek and Pavel Petkov. 2006. *Zweitsprachige Lexikographie und Deutsch als Fremdsprache*: 49-58. Hildesheim: Georg Olms.
- Gouws, Rufus H. en Danie J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Hanks, P.** 2004. *Word Lists for Big Dictionaries*. Message posted to the Lexicography List, 15 March 2004. <http://groups.yahoo.com/group/lexicographylist/message/505>.
- Keyser, Helane.** 2003. *Ondersoek na 'n model vir die opleiding van leksikograwe vir verklarende woordeboeke*. Ongepubliseerde M.Phil.-tesis. Stellenbosch: Universiteit van Stellenbosch.
- Lombard, F.J.** 1992. Voorbeeldmateriaal in woordeboeke. *Lexikos* 2: 148-164.
- Martins, Magda.** 2005. WAT. Word alles in dié woordeboek toegelaat? *Beeld*, 5 September.
- Odendal, François F. en Rufus H. Gouws.** 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Kaapstad: Pearson.
- Swart, Corlietha.** 2006. WAT vier 80 met 12de deel. *Volksblad*, 27 Februarie.
- Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns.** 2002. *Afrikaanse woordelys en spelreëls*. Kaapstad: Pharos.
- Van den Berg, Leon.** 2006. Veel meer as 'n woordeboek. *Beeld*, 15 Mei.
- Wiegand, H.E.** 1996. Über Mediostrukturen bei gedruckten Wörterbüchern. Zettersten, A. en H. Pedersen (Reds.). 1996. *Symposium on Lexicography VII*: 11-43. Tübingen: Max Niemeyer.

The African Languages Research Institute: A Milestone in the Development of the Zimbabwean Languages*

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

Abstract: This article is an assessment of the work the African Languages Research Institute (ALRI) has done towards developing the indigenous languages of Zimbabwe. It looks at what the research team at ALRI has achieved, first in initiating serious research on the Zimbabwean languages and the progress it has made towards achieving its goal, developing and raising the status of these languages. It also considers what ALRI has planned for the future of the different categories or levels into which the various Zimbabwean languages have been classified. Part of the assessment focuses on measures ALRI has put in place to ensure the initiated research programme is sustainable and will continue in future. The article furthermore discusses the importance of the research work being done at ALRI, especially with regard to language development as a means towards self-realisation and -actualisation, national advancement and the sustenance of the languages involved. ALRI's agenda is also analysed to see how well it agrees with popular thinking in Zimbabwe concerning the development and promotion of all of the indigenous languages.

Keywords: ALRI, CAPACITY BUILDING, INSTITUTIONALISATION, LANGUAGE HARMONISATION, LANGUAGE STANDARDISATION, MONOLINGUAL LEXICOGRAPHY, ZIMBABWEAN LANGUAGES

Opsomming: Die African Languages Research Institute: 'n Mylpaal in die ontwikkeling van die Zimbabwiese tale. Hierdie artikel is 'n evaluering van die werk wat die African Languages Research Institute (ALRI) gedoen het in die ontwikkeling van die inheemse tale van Zimbabwe. Dit kyk na wat die navorsingspan bereik het, eerstens deur ernstige navorsing oor die Zimbabwiese tale te onderneem en die vordering wat gemaak is met die bereiking van sy doel, die ontwikkeling en die statusverhoging van hierdie tale. Dit beskou ook wat ALRI beplan het vir die toekoms van die verskillende kategorieë of vlakke waarin die onderskeie Zimbabwiese tale geklassifiseer is. 'n Deel van die evaluering fokus op maatreëls wat ALRI in plek gestel het om te verseker dat die onderneemde navorsingsprogram volhoubaar is en in die toekoms sal voortgaan. Verder bespreek die artikel die belangrikheid van die navorsingswerk wat by ALRI gedoen word, veral met betrekking tot taalontwikkeling as 'n manier tot selfverwesenliking en -aktualisering.

* This article is based on a paper presented at the Sixth International Conference of the African Association for Lexicography organised by the Sesotho sa Leboa National Lexicography Unit, University of the North (at present University of Limpopo), Polokwane, Republic of South Africa, 8–10 July 2001.

nasionale vooruitgang en die onderhoubaarheid van die betrokke tale. ALRI se agenda word ook ontleed om te sien hoe goed dit ooreenstem met die algemene denke in Zimbabwe betreffende die ontwikkeling en bevordering van al die inheemse tale.

Sleutelwoorde: ALRI, VERMOËNSBOU, INSTITUSIONALISERING, TAALHARMONISERING, TAALSTANDAARDISERING, EENTALIGE LEKSIKOGRAFIE, ZIMBABWIESE TALE

1. Introduction

In this article, which focuses on the contribution of the African Languages Research Institute (ALRI) towards the development of Zimbabwe's indigenous languages, language development is viewed as involving two steps. The first is that of reducing a language to writing, the aim being that of standardising it to become a literary language. In this case, the standardisation process results in the establishment of an alphabet for the language and the rules of spelling and word division to be followed when writing that language. The rules are generally referred to as the language's orthography. The second step, which follows after the establishment of a language's orthography, is the development of tools that further enhance the research on and documentation of a particular language. Some such tools include general and specialised dictionaries, corpora, morphological parsers, syntactic analysers and spell checkers, the compilation of which rely on the agreed or standard orthography of a language. In the light of what is involved in these two steps, it becomes clear that language development is an ongoing process of researching and updating a language's orthography and its descriptive tools. Once a standard orthography has been developed and the language has been adequately described and documented, its status is automatically raised. In this article, it will be considered how ALRI is trying to raise the status of Zimbabwean indigenous languages through its research agenda.

To date, efforts towards developing the African languages spoken in Zimbabwe are more than a century old. Researchers, both foreign and local, as individuals or small groups, have made varying contributions towards the standardisation and documentation of the country's different languages. The earliest efforts comprised missionary activities, which mainly centred on the development of orthographies for various languages and/or dialect clusters. This was realised through different missionary groups strategically located in different parts of the country where they set up presses for printing Bibles, church literature and educational books to create literature for their new converts. Although missionary efforts have often been viewed as aimed at creating a literate group of individuals to use in spreading the gospel, the local languages ended up gaining in the process by having their orthographies developed. Examples where missionary activities benefited local languages can be drawn from developments in Shona, Kalanga and Nambya. For Shona, for example, missionary activities based in different parts of Mashonaland led to

the creation of dialect clusters, identified as Zezuru, Karanga, Manyika, Ndau and Korekore. The missionaries working on these particular dialect clusters were then instrumental in developing orthographies for each of them. However, after the establishment of a central colonial government and also following the general problems of writing differently in rather highly mutually intelligible dialects, a standard Shona orthography was created through the recommendations of Doke (1931) in his *Report on the Unification of the Shona Dialects*. Since then, the Shona writing system is largely based on these recommendations. Doke's writing system was successful because all the background research and necessary documentation had been done by missionaries, who had championed the idea of promoting literary Shona. It was also the missionaries who gave the necessary personal and financial assistance to Doke. Similar missionary efforts were witnessed at Inyati mission in Plumtree, which led to the development of a writing system for Kalanga. A writing system for Nambya again resulted from efforts by the Catholic missionaries based at St Mary's mission in Hwange.

Following the establishment of writing systems in the different languages through missionary activities, researches on Zimbabwe's indigenous languages, by individuals and groups of linguists and non-linguists flourished. Basic grammar books were produced, especially for Shona and Ndebele which had been selected as the indigenous languages to be taught in schools in Mashonaland and Matebeleland respectively (see Doke 1931). Bilingual dictionaries were also published (see, for example, Hannan 1959, Dale 1981, and Moreno 1988, among others). Whilst all these efforts should be acknowledged for pioneering the research on and documentation of the various indigenous languages, they were either limited in scope or lacked the much needed focus towards the development and raising of the status of these languages. For example, most of the work on documenting the indigenous languages ensued from research by individuals towards the attainment of their academic qualifications at Honours, Masters and Doctoral levels. Whilst these are valued works, they tended to be narrowly focused on specific linguistic aspects which lacked the necessary impact or influence needed in the formulation of national policies to be followed in the development and raising the status of the Zimbabwean languages. As has also been observed, the dictionaries published were bilingual. Generally speaking, bilingual dictionaries are mainly meant for second-language speakers. In the case of the Zimbabwean languages, the dictionaries were compiled for the use of the colonial authorities. Such dictionaries do not effect as direct an influence on language development as monolingual dictionaries, neither do they contribute to empower the concerned languages.

Language researches carried out before the establishment of ALRI were not coordinated. The studies tended to narrowly focus on one language or dialect or another and also to be based on short-term research programmes. However, most language researchers in Zimbabwe, just as in other countries of the world, have now realised that their efforts can bear more fruit if they are coor-

minated at local, regional, national and even international levels. They have also realised that through such coordination larger and long-term research programmes can be embarked on in a more sustainable manner. Such a coordination of research programmes for the Zimbabwean languages has come with the establishment of ALRI.

2. What is ALRI?

ALRI is a non-faculty, interdisciplinary research unit based at the University of Zimbabwe, and is dedicated to the development of all indigenous languages spoken in Zimbabwe. Its goals are contained in its mission statement, which is to research, document and develop the Zimbabwean indigenous languages in order to promote and expand their use in all spheres of life. Formerly the Institute was the African Languages Lexical (ALLEX) Project¹ housed in the Department of African Languages and Literature at the University of Zimbabwe. The ALLEX Project, which was launched in 1992, had the compilation of a variety of general and specialised monolingual dictionaries in Zimbabwe's two national languages, Shona and Ndebele, as its main goal. Thus, the work of the ALLEX Project (1992–2000) has culminated in the publication of two general medium-sized Shona monolingual dictionaries, *Duramazwi reChiShona* (DRC) (Chimhundu 1996) and *Duramazwi Guru reChiShona* (DGC) (Chimhundu 2001) as well as a general medium-sized Ndebele monolingual dictionary, *Isichazamazwi SesiNdebele* (ISN) (Hadebe 2001). It has also seen the building and maintenance of fairly large corpora for Shona and Ndebele which have since been used to aid the process of compiling the ALLEX dictionaries. The successful dictionary making has also allowed for work to be commenced on the development of other language research tools such as morphological parsers and syntactic analysers for the two languages.

ALRI was established in 2000, transforming the ALLEX Project into a permanent research unit. As noted in Chimhundu (2005: 2), the decision to create ALRI resulted from the intention to institutionalise the experimental work done by the ALLEX Project so that the scope of its research agenda could be expanded and broadened to include more local languages. The success the Project had achieved in its experimental research on and documentation of Shona and Ndebele was taken as a model in the programmes for the other local languages.

3. ALRI's research activities

As a means towards achieving its goals of developing and raising the status of the country's languages, ALRI has embarked on a number of research activities with each activity aimed at a particular aspect of language development. The next sections will discuss some of the activities with which the ALRI research teams have commenced. These are described under the following subheadings.

3.1 Language standardisation

As already noted, standardisation is the first step towards the development of any language. It is only when standardised that a language can further be developed so as to be able to function in all spheres of life. It is also when a language has been standardised that it can be used as an efficient communicative tool in, for instance, education, government, industry, commerce and law. It is because of its importance that ALRI has put the standardisation of all of Zimbabwe's languages at the core of its research agenda. In Zimbabwe, some languages already have standard orthographies while others have no writing systems at all. Because of this, ALRI has defined its role as twofold. With regard to those languages not yet standardised, its role is that of developing common writing systems for them. The bulk of the languages without standard orthographies are the country's 'minority' or community languages, examples of which include, among others, Shangani, Tonga, Sotho, Chewa, Tswawo and Venda. The planned work at ALRI is such that once standard writing systems have been developed for these languages, basic grammars can then be written. Subsequently, it would become possible also to teach these languages as subjects in schools. The languages can further be developed through modernisation.

With regard to languages already having orthographies, the standardisation process will be advanced by the revision and refinement of the existing orthographies as well as the writing of new and the updating of existing grammars, basing decisions on new knowledge gained from new research methods. For example, research advances so far made in Shona corpus-assisted lexicography have shown the need for the inclusion into the Shona alphabet of letters and clusters not acceptable in the current Shona orthography. Some such letters and clusters include *l*, *th*, *ml* and *dhl*. The letter *l*, which is often introduced into the language through borrowed words from English, has, for example, been discarded from the Shona orthography on the assumption that wherever it is encountered in borrowed words, it will consistently be replaced by *r*. However, corpus research has shown the increased use of this letter by Shona speakers in their daily lives, hence the need to revise the current Shona orthography in order to include the *l*. The same can be said of the *th* cluster which had also been discarded on the assumption that it would be replaced in Shona by the letter *t*. In Ndebele, the increased use of the letter *r*, which is not part of the current Ndebele orthography, has also necessitated the need for revising Ndebele's orthography to make the inclusion of *r* possible.

3.2 Monolingual lexicography

As already intimated, the major objective of both the ALLEX Project and ALRI is to develop tools enhancing research on and the development of the indigenous languages of Zimbabwe. Such tools include monolingual dictionaries,

whose importance in language standardisation and documentation need not be over-emphasised. They play an important role in developing a language for it to be used in all aspects of life. For example, their compilation usually leads to the development of terminology employed in specialised fields, which later becomes the basis for expanding the language concerned. The reason for this is that dictionaries, especially monolingual dictionaries, have a standard-setting influence since they are generally used by language speakers as their main tools in interpreting the rules about the spelling, meaning and pronunciation of words.

Monolingual lexicography only started with the ALLEX Project under whose leadership, funding and guidance the publication of the first-ever monolingual dictionaries in Shona and Ndebele have been made possible. The production of general and specialised monolingual dictionaries has had a positive impact on the process of raising the status of Shona and Ndebele. Mention of a few Shona monolingual dictionaries published or being compiled under the ALLEX/ALRI research programme would suffice as illustrative examples of how monolingual lexicography has led to the recognition of Shona as a language that can be used in any sphere of life, including specialised areas. To start with, the publication of DRC and DGC had some impact on the general public and professionals in different specialised fields. The result was that professionals from various fields approached ALRI for collaborative research aimed at compiling dictionaries, glossaries and other reference works for their respective fields of expertise. One such work that resulted from this collaborative research is a small-to-medium sized monolingual dictionary of Shona biomedical terms, *Duramazwi reUrapu neUtano* (Mpofu et al. 2004), which was a product of a collaborative research venture between researchers from ALRI and others from the Institute of Continuing Health Education (ICHE). Other similar research projects currently under way include one on a dictionary of Mathematics terms and another on a dictionary of Physics terms, both of which are being carried out by researchers at ALRI in collaboration with specialists in the respective fields of study. These are projects the respective specialists brought to ALRI for assistance, and are being carried out in addition to what ALRI itself had planned. Other projects still at the planning stage include one for a dictionary of agricultural terms and another for a dictionary of geographical terms. This kind of response is evidence of the extent of the influence ALRI's work has had on the Zimbabwean society at large.

Other ALRI projects pertaining to terminological dictionaries include the Shona dictionary of musical terms, *Duramazwi reMimhanzi* (Mheta et al. 2005). Also completed are two specialised dictionaries, the Ndebele dictionary of musical terms *Isichazamazwi SezoMculo* (Nkomo and Moyo 2006) and the Shona dictionary of linguistic and literary terms *Duramazwi reDudziramutauro neUvaranomwe* (Chimhundu and Chabata 2007). Work is also under way on, among others, the revised and enlarged Ndebele dictionary. Planned work at ALRI furthermore include monolingual lexicographical projects in Zimbabwe's

minority languages, starting with those that have been prioritised by both ALRI and the Ministry of Education, that is, Kalanga, Nambya, Shangani and Tonga. Through general and terminological dictionaries, technical terms for different specialist fields are collected or coined and recorded in systematic ways helpful to speakers or users of a particular language. The availability and standardisation of these terms through monolingual dictionaries will eventually make it possible to use the indigenous languages as communicative tools in all aspects of life. This will, in turn, raise their status for it would become possible for them to be used in important areas such as education, government, industry, commerce and law.

Monolingual dictionaries are also compiled as a logical step towards language modernisation. With the help of decisions based on language descriptions contained in monolingual Shona and Ndebele dictionaries, morphological parsers, syntactic analysers and spell checkers are being developed for Shona and Ndebele. These research tools make automatic language analysis possible.

3.3 Corpora building and maintenance

A corpus has been defined by Renouf (1987: 1) as 'a collection of texts, of the written or spoken word, which is stored and processed on computer for the purposes of linguistic research'. According to this description, it implies an electronic or machine-readable body of texts. Reference is made to a corpus stored in the computer from which words, phrases and sentences or even chunks of text can easily be retrieved. The importance of electronic corpora as valuable sources of evidence for linguistic and other language related analyses need not be over-emphasised. As observed by Kennedy (1998: 88), for example, a linguistic corpus, in whatever form, is important as a basis for more accurate and reliable descriptions of how languages are structured and used. Corpora are very important tools in providing conclusive evidence for researches in fields such as lexicography, dialectology, semantics, syntax, psycholinguistics, phonology, sociolinguistics, code-switching and code-mixing. These are disciplines whose focus is on the manner of how language is either structured or used, or in which evidence for any controversial issues can only come from instances where language is found in use. It is because of the importance of this resource that ALRI has made the building and maintenance of corpora in the various indigenous Zimbabwean languages one of its priority research activities. This is especially important because the research done by ALRI is based not only on theoretical linguistics, but also on linguistic data showing how speakers use language. The aim is to publish reference works, which reflect and represent actual, not theoretically possible language usage.

To date, fairly large corpora have been compiled for Shona² and Ndebele, the two main indigenous languages of Zimbabwe. The materials for the corpora come from both transcribed oral interviews on various aspects of life and written literature of various kinds, which are later encoded or scanned, tagged

and parsed before they are stored as text and sound in electronic form. In addition, there is also an image corpus being built for the benefit of dictionaries that would need illustrative examples in the form of pictures. The images are drawn and stored in a database for future use when required. Work on building corpora in some of the country's minority languages have since started, with serious work already commenced on Kalanga, Nambya, Tonga and Shangani. The corpora in these community languages are modelled according to the standards adopted in the building of the Shona and Ndebele corpora. Although the ALRI corpora are primarily constructed for lexicographic use, they have been found to be more and more useful in other language and language-related studies already identified above as dialectology, semantics, syntax, psycholinguistics, phonology, sociolinguistics, code-switching and code-mixing. In these studies, text corpora, for example, are important in providing contexts, both linguistic and cultural, in which words and phrases are used. This is especially possible since part of the corpora comprise texts from oral interviews taken from naturally occurring language in use by mother-tongue speakers of the respective languages. Currently, ALRI text corpora, which are also accessible to international researchers on the Internet, are being exploited for various language and language-related activities the world over. For example, the Shona, Ndebele and Nambya corpora are extensively used by Zimbabwean researchers studying abroad as well as at institutions in other parts of Africa.

3.4 Language harmonisation

Language relatedness is one of the fascinating aspects of any linguistic study. In the case of Africa and Zimbabwe, most indigenous languages are generally related by sharing the same morphological, phonological and syntactic systems. Unfortunately the history of colonisation has led to the creation of artificial boundaries between these otherwise highly related language varieties. The missionaries and colonial administrators who established programmes for the development of orthographies for indigenous languages deliberately ignored the similarities between language varieties and decided to develop divergent orthographies even for mutually highly intelligible varieties. As noted by Simango (2003: 33), the use of diverse orthographies for mutually intelligible language varieties is detrimental to the development of the concerned languages for two basic reasons. Firstly, it reduces the level of mutual intelligibility, through the written word, among language varieties that were and still remain mutually intelligible in their spoken forms. For example, it makes it very difficult for communities who use diverse orthographies to share information conveyed via the written medium. Secondly, divergent orthographies tend to distort the ground-level realities by exaggerating the differences existing among language varieties, at the same time obscuring their similarities. Sadly for Zimbabwe, this historical fact has also led to the elevation of some dialects to the status of independent languages and in the process entrenching

divisions among groups of people who were previously and who can still easily be united by the same language.

Because of the disadvantages of using diverging orthographies for mutually intelligible language varieties, research efforts in Sub-Saharan Africa are now concentrated on harmonising the orthographies for related languages. There has been a realisation that people can be better served if orthographies are harmonised in such a way that the written forms reflect the similarities existing between varieties of each language cluster. In Zimbabwe, such a programme will be realised through efforts by ALRI. Language harmonisation has been considered important for ALRI with the realisation that the bulk of Zimbabwe's language varieties (which have since been officially recognised as independent languages and have had different orthographies developed for them) are mutually highly intelligible in speech. Despite the fact that different orthographies have been designed for them, Shona, Kalanga and Nambya, for example, are to such an extent mutually intelligible that speakers from the three languages can easily communicate with very few difficulties. It is because of this high mutual intelligibility that Kalanga has often been described as one of the Shona dialects, and also that Nambya is sometimes regarded a subdialect of Kalanga. ALRI's philosophy behind harmonising such related languages is that it will foster greater unity amongst linguistically related peoples. This is also hoped to widen readership and markets for written materials. In other words, it will become easier and cheaper to produce one document that can be circulated widely than to reproduce the same document in many different orthographies of languages otherwise closely related.

ALRI's planned harmonisation programme will be at two levels. Firstly, it will try to harmonise related language clusters. For example, attempts will be made to design a common orthography for Eastern Shona (Shona spoken in the eastern parts of Zimbabwe and Mozambique), Central Shona (which is spoken in many parts of Zimbabwe) and Western Shona (spoken in the western parts of Zimbabwe and Botswana and generally referred to as Kalanga). Secondly, the programme will see the harmonisation of languages Zimbabwe shares with its neighbours defined as separate languages only on the basis of national boundaries, but otherwise having high mutual intelligibility. Examples of language clusters in need of harmonisation include South African and Zimbabwean Ndebele, Mozambican and Zimbabwean Shona, Botswanan and Zimbabwean Kalanga, Zambian and Zimbabwean Tonga and Zimbabwean, Mozambican and South African Shangani.

4. Strategies for development

All the different research activities discussed above are aimed at developing Zimbabwean languages for use in all aspects of life and also at modernising them in the process. However, the activities can only be carried out successfully within the framework of a well-designed and sustained research programme. It is in this spirit that ALRI has adopted a number of strategies con-

sidered necessary for undertaking a sustained research programme. In this section, a few of these strategies will be discussed.

4.1 Institutionalisation

One of the strategies adopted by ALRI is the institutionalisation of the work initiated by the ALLEX Project through the creation of a permanent research institute. With the realisation that the ALLEX Project will, like any other project, have a life after which it would be concluded, institutionalisation was seen as a way of ensuring long-term planning. Institutionalisation would mean much for the sustained development of Zimbabwean languages. Firstly, institutionalisation entailed that lexicography and the other research activities initiated under the ALLEX Project would become permanent activities without a time limit. Secondly, it also meant that the responsibility of supporting the research agenda would become national and local. This kind of support lasts longer when compared to donor-funded programmes only continuing for the period donors are willing to provide support. As fields like lexicography and corpora building are multidisciplinary, unnecessary constraints are placed on any such activity housed within a department with its own needs and goals which guide its development as a discipline. The transformation from a project within a department to an autonomous research unit has, for example, made it possible and relatively easy for researchers at ALRI to organise collaborative projects with researchers from other institutions and other fields of study.

4.2 Capacity building

It has already been noted that research on Zimbabwean languages has had a long history. However, the greatest part of this history has been dominated by individual researchers who often were non mother-tongue speakers of the respective languages, and who also tended to work on short-term research projects. For example, it has already been observed that research on the country's indigenous languages started with missionaries who studied and documented the languages and dialects for evangelising purposes. In most of these projects, mother-tongue speakers of respective languages were only valuable as research informants or assistants during the data collection stage. The speakers of the different languages were not trained to become experts in linguistic research. What this entailed was that at the end of each programme no mother-tongue speakers of the respective languages were equipped with the knowledge of continuing with or improving on the work initiated. This is rather different from the approach adopted at ALRI where great emphasis has been laid on capacity building. From the outset, a training component has been built into the ALLEX/ALRI programme. Through its training programme, members of the Institute, who are all mother-tongue speakers of one or another of Zim-

babwe's indigenous languages, are trained in all aspects of research activities taking place at the Institute. Researchers in Shona and Ndebele, who received their training and experience from their research collaborators from Scandinavia through the ALLEX Project, have had the responsibility of training other speakers of these and other Zimbabwean languages in lexicographic and other language-related activities such as corpus building and maintenance, dialectology and others. The idea is that these speakers, once trained, would spearhead the research and documentation of their respective languages. This entails that the undertaking of documenting and developing Zimbabwean languages will not end with ALRI, but will be continued.

The pool of necessary expertise is being created through two kinds of training. There is the element of formal academic training at Honours, Masters and Doctoral levels through the provision of scholarships to members to study in the fields of linguistics and lexicography. One of the requirements for taking up the scholarships is that the recipients would develop their dissertations or theses around some linguistic or lexicographic aspect of their mother-tongues. This requirement is meant especially to initiate research on and documentation of the country's minority languages. The idea is to move away from the missionary approach of involving mother-tongue speakers only as research informants, cultivating a sense of ownership of the research agenda amongst mother-tongue participants. This approach has been adopted for two basic reasons. Firstly, it lays the foundation for sustained development. Secondly, researchers at ALRI cannot accomplish all the required work on all of the country's languages by themselves. Because of this, there is a need to train mother-tongue speakers of different languages who are expected to contribute substantially towards research programmes in their respective language communities. In such cases, ALRI's role would only become advisory and supervisory.

Training has also taken the form of practical on-the-job training, especially with regard to new computational language research methods. This kind of training is done through annual workshops organised for members. These have been highly successful as practical training for various IT linguistic and lexicographic tasks and also for project planning, budgeting and management. Through linkage with other institutions of higher learning which have a language component, capacity building has been extended even to non-ALRI staff members, the idea being that expanding the capacity building programme would create a culture of collaborative research, which ensures sustainability of quality research at national level.

4.3 Community involvement

Any kind of research involving humans needs the community, which is both the object of study and the principal beneficiary of the results. For the research programme at ALRI, community involvement is an indispensable component. Apart from providing the much needed moral and logistical support and the

required data, the community is also the consumer of ALRI's products. The idea, therefore, is that the community should understand and also be actively involved in the research programme. In this way, the community is afforded an opportunity to decide what is of immediate importance to them; thus also participating in determining the course the research programme should follow. Only when this is achieved, the programme can be appreciated, supported and also be assured of success and continuity.

Since its inception ALRI has made its research agenda known to the Zimbabwean community through its various activities. Some such activities include its outreach programmes, which over the years have been carried out in all the districts across the country. During these programmes, those with an interest in language research are brought together to have formal and informal discussions with researchers from ALRI and to make their contributions towards the research agenda. The community has also been engaged in the Institute's field researches. Because of this, the Zimbabwean community now feels part of the ALLEX/ALRI research programme, showing its willingness to actively participate in the research activities initiated by the Institute.

4.4 Collaborative research

Unlike the preceding research programmes which were either conducted by individuals or small groups of researchers aimed at investigating a particular aspect of language, ALRI's programme is much bigger, incorporating specialists in different but relevant fields of study. It is a collaborative research programme involving computer scientists, lexicographers, linguists, teachers, the community of speakers and other relevant stakeholders. For example, specialised dictionary projects carried out so far have seen the active involvement of and representation from all universities and teachers' colleges which offer training in the African languages of Zimbabwe, and some education officers and university students acting as research assistants during data collection and other fieldwork activities.

In the research programmes carried out before, ALRI has also tended to compartmentalise some languages by treating them as if they were only spoken in Zimbabwe. Researches on cross-border languages, for example, have not recognised the fact that these languages are also being studied in neighbouring countries. With regard to such languages, the programme at ALRI will see a collaborative approach involving researchers and research institutes from neighbouring countries co-operating for the common goal of developing the languages concerned. This means that research on Kalanga, for example, will involve researchers from Zimbabwe and Botswana, Shangani researchers from Zimbabwe, South Africa and Mozambique, and Tonga Zimbabwean and Zambian researchers. A collaborative research has two advantages. Firstly, scholars from different institutions and countries will be afforded a chance to share research ideas, experiences and findings. Secondly, research costs are also shared. This sharing of human and financial resources naturally leads to a fruit-

ful programme where quality results are achieved quickly and cheaply. Such research links have already been initiated and logistics are currently being worked out for collaborative research to commence.

5. Conclusion

In this article, a description has been given of the work ALRI is doing towards developing and raising the status of the indigenous languages of Zimbabwe. Some of the strategies ALRI has put in place as measures of ensuring that its research programme is sustained in future have also been discussed. The discussion also aimed to highlight some of the reasons why the status of Zimbabwe's indigenous languages could not be raised despite the fact that research on them has had a long history. However, the success ALRI has achieved in cultivating a love for the indigenous languages has been shown as evidence for the use of its monolingual dictionaries by people who traditionally tended to look down upon these languages.

Notes

1. Currently, the ALLEX Project which receives the bulk of its funding from NUFU, the Norwegian Programme for Development, Research and Education, is the main project being undertaken at ALRI. Through this project, Zimbabwean scholars obtained training in computational lexicography and in corpora building and maintenance from their counterparts in Norway and Sweden through collaborative research, mainly with scholars and/or researchers at the Universities of Oslo (Norway) and Gothenburg (Sweden).
2. For detailed discussions on how the Shona corpus, for example, was compiled, refer to Chabata (2000) and Mberi (2002).

References

- Chabata, E. 2000. The Shona Corpus and the Problem of Tagging. *Lexikos* 10: 75-85.
- Chimhundu, H. (Ed.). 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.). 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Chimhundu, H. 2005. *Lexicography and Language Raising: Dictionaries in Zimbabwean Languages*. Public lecture presented as part of the Golden Jubilee Celebrations of the University of Zimbabwe, Harare, 13 April 2005.
- Chimhundu, H. and E. Chabata (Eds.). 2007. *Duramazwi reDudziramutauro neUvaranomwe*. Gweru: Mambo Press.
- Dale, D. 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.
- Doke, C.M. 1931. *Report on the Unification of the Shona Dialects Carried Out under the Auspices of the Government of Southern Rhodesia and the Carnegie Corporation*. Hertford: Stephen Austin.
- Hadebe, S. (Ed.). 2001. *Isichazamazwi SesinDebele*. Harare: College Press.
- Hannan, M. 1959. *Standard Shona Dictionary*. Salisbury: The College Press.

- Kennedy, G.** 1998. *An Introduction to Corpus Linguistics*. London/New York: Longman.
- Mberi, N.E.** 2002. *The Categorical Status and Functions of Auxiliaries in Shona*. Unpublished D.Phil. Thesis. Harare: University of Zimbabwe.
- Mheta, G. et al. (Eds.)**. 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Moreno, A.** 1988. *Nambya Dictionary*. Gweru: Mambo Press.
- Mpofu, N. et al.** 2004. *Duramazwi reUrapu neUtano*. Gweru: Mambo Press.
- Nkomo, D. and N. Moyo (Eds.)**. 2006. *Isichazamazwi SezoMculo*. Gweru: Mambo Press.
- Renouf, A.** 1987. Corpus Development. 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*: 1-40. London: Collins.
- Simango, S.R.** 2003. Do We Dare Harmonise Orthographies? Chebanne, A. et al. (Eds.). *Unifying Southern African Languages: Harmonisation and Standardisation*. 2003. Cape Town: Centre for Advanced Studies of African Society (CASAS).

Language Planning and Monolingual Dictionaries: With Special Reference to Ndebele*

Samukele Hadebe, *Department of African Languages and Literature, University of Zimbabwe, Harare, Zimbabwe (shadebe@arts.uz.ac.zw)*

Abstract: The first monolingual Ndebele dictionary, *Isichazamazwi SesiNdebele*, had a number of effects on Ndebele, some of which with implications for language planning. One such language planning activity was the standardization of Ndebele. The article focuses on the standardization of vocabulary and spelling. Lexicographers and most of those interested in lexicographic issues are familiar with the challenges posed by what constitutes the standard vocabulary or the standard meaning of words. These questions were crucial for a general monolingual dictionary like *Isichazamazwi SesiNdebele*. General dictionaries are the standard dictionaries for particular languages, assumed to be reflective of the 'standard usage' of that given language in terms not only of spelling but also of meaning.

The Ndebele dictionary is based on a corpus which means that words perceived by some as foreign or as 'bad' language are considered for lemmatization. Problems were also encountered with the spelling of these loanwords. By making decisions on which words to lemmatize and how to spell loanwords, lexicographers become involved in language planning matters. The article draws from the Ndebele dictionary-making experience to discuss the role of monolingual African language dictionaries in language planning in general.

Keywords: STANDARDIZATION, GENERAL MONOLINGUAL DICTIONARY, NDEBELE, LOANWORDS, LANGUAGE PLANNING, SPELLING, STANDARD, VOCABULARY, STATUS PLANNING, CORPUS PLANNING

Opsomming: Taalbeplanning en eentalige woordeboeke: Met spesiale verwysing na Ndebele. Die eerste eentalige Ndebelewoordeboek, *Isichazamazwi SesiNdebele*, het 'n aantal gevolge vir Ndebele gehad, sommige waarvan met implikasies vir taalbeplanning. Een so 'n taalbeplanningsaktiwiteit was die standaardisering van Ndebele. Die artikel fokus op die standaardisering van woordeskat en spelling. Leksikograwe en meeste van diegene wat hulle in leksikografiese kwessies interesseer, is bekend met die uitdagings gestel deur wat die standaardwoordeskat of die standaardbetekenis van woorde behels. Hierdie vrae is beslissend vir 'n algemene eentalige woordeboek soos *Isichazamazwi SesiNdebele*. Algemene woordeboeke is standaardwoordeboeke vir bepaalde tale, wat aanvaar word om die "standaardgebruik" van daardie taal te weerpieël, nie net wat die spelling nie, maar ook wat die betekenis betref.

* This article is based on a paper presented at the Tenth International Conference of the African Association for Lexicography, organized by the Sesiu sa Sesotho National Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

Die Ndebelewoordeboek is op 'n korpus gebaseer, wat beteken dat woorde wat deur sommige as vreemde of as "swak" taal beskou word vir lemmatisering oorweeg word. Probleme met die spelling van hierdie leenwoorde is ook teëgekem. Deur besluite te neem oor watter woorde om te lemmatiseer en hoe om leenwoorde te spel, raak leksikograwe betrokke by taalbeplanningsaangeleenthede. Die artikel steun op ervaring verkry uit die maak van 'n Ndebelewoordeboek om die rol van eentalige Afrikataalwoordeboeke in taalbeplanning in die algemeen te bespreek.

Sleutelwoorde: STANDAARDISERING, ALGEMENE EENTALIGE WOORDEBOEK, NDEBELE, LEENWOORDE, TAALBEPLANNING, SPELLING, STANDAARD, WOORDESKAT, STATUSBEPLANNING, KORPUSBEPLANNING

1. Introduction

The compilation of the first monolingual Ndebele dictionary, *Isichazamazwi SesiNdebele*, henceforth ISN, resulted in a number of language planning activities.¹ The role of dictionaries in standardizing language has often been established (Chimhundu 1997, Drame 2001, Hadebe 2006, Wolff 2000). Standardization, an essential feature of language planning in general, is probably the most significant corpus planning activity carried out on Ndebele for dictionary compilation purposes. Lexicographers and most of those interested in dictionary issues are familiar with challenges posed by what constitute the standard vocabulary or the standard meaning of words. Monolingual general-purpose dictionaries like the ISN are the standard dictionaries for their respective languages and are assumed to reflect the 'standard usage' of that language in terms not only of spelling but also of meaning. Writing on the relationship of dictionaries to standard language, Alberts (2005: xiv) says: "Lexicographers compile dictionaries by (usually) documenting words in the vocabulary of a standard language. Should they compile a dictionary that does not concern the standard variety of a language it is usually called a dialectal dictionary." This observation is of significance for languages like Ndebele that could hardly be said to have had a standard prior to the production of the first monolingual dictionary and a grammar book.

In this article, it is shown how the ISN has an impact on language planning by presenting an overview of the notion of language planning, discussing the significance of monolingual dictionaries, outlining the sociolinguistic situation of Ndebele and sketching the language planning activities resulting from the production of the ISN.

2. What is Language Planning?

Language planning as a concept is problematic not only because of the different meanings attached to it but also because the possibility of planning a language is questionable. Partly for this reason, there has so far been no consensus among scholars on a definition of language planning. This is reflected in the

great variety of terms that exist to describe language planning. These include the terms *linguistic reform*, *language reform*, *deliberate language change*, *planned language change* and *language treatment* as well as terms such as *language engineering* (Pauwels 1998: 2) and lately *language management* (Ntshangase 1997: 17).

Language planning has been defined as "an explicit choice among (linguistic) alternatives" (Fasold 1984: 246), or as "a deliberate language change ... characterized by the formulation and evaluation of alternatives for solving language problems to find the best (or optimal, most efficient) decision" (Rubin and Jernudd 1971: xvi). Jernudd and Das Gupta (1971: 211) define it as "an orderly decision-making about language on a national level". Mkanganwi (1992: 6) sees it as "the conscious, predictive approach to language and language use". From these views on language planning, some of the key concerns of language planning may be drawn. Firstly, the issue of language diversity and hence the choice among many alternatives may be noted. Secondly, language planning is not accidental but intentional, hence the terms 'conscious', 'orderly decision-making' and 'deliberate'. It then introduces the question of who does language planning and why. Language planning is always directed towards a certain goal in order to solve perceived language problems. Language planners are "the wielders of political-economic power in a state, nation, or nationalising entity" (Eastman 1992: 96). Jernudd and Das Gupta (1971: 196) say: "The broadest authorization for planning is obtained from the politicians. A body of experts is then specifically delegated the task of preparing a plan."

The view that language authorization comes from a central authority has been challenged by some scholars like Bamgbose (1991) and Alexander (1992) as there are numerous players in language planning. Some of these are non-governmental, such as publishers, the media, churches and individuals. It should also be noted that language planning is a very broad activity involving both political and linguistic decisions. The more political aspects seem to be those relating to status planning while the linguistic aspects involve corpus planning activities. For this reason, it can be said that language planning includes all the activities ranging from selection and delegation of functions to each variety to codification, fixing grammar and pronunciation rules as well as elaborating the various features of the language and the implementation of these plans (Hadebe 2006). Unlike status planning, corpus planning is usually an on-going process, which can be carried out in anticipation of the possible functions the language might have in future or may be performed in order for the language to cope with functions it is already serving. The production of the ISN had a more significant impact on corpus planning than on status planning as will be shown in this article. To be more specific, the dictionary had a standardizing influence on Ndebele.

3. Monolingual Dictionaries in Society

Monolingual general-purpose dictionaries are mostly assumed to represent

and reflect the lexicon of the language in question and hence the richness of its culture. Therefore, these dictionaries are seen not only as linguistic tools but also as repositories of a particular people's culture. Béjoint (2000: 137-138) writes:

General-purpose dictionaries occupy a very special position in all the societies that produce them ... The emblematic power of the general-purpose dictionary is so strong, so real in a way, that the dictionary is felt to be necessary to any nation that wants to be recognized as an independent entity. It is one of the rare objects that can materialize the existence of a language, and hence of a nation, acting as a symbol of the unification of a community.

This cannot be nearer the truth regarding the impact of the ISN on the Ndebele. Being the first monolingual dictionary and being the first research work of its kind, the ISN received nation-wide publicity and caused public excitement after its appearance. It can safely be claimed that the dictionary will be the standard reference work for the Ndebele language for quite some time (Hadebe 2006: 29). The Ministry of Education and Culture actually prescribed the ISN as a reference book in schools. In the prestigious National Arts Merits Awards (NAMA) of 2002, it won a first prize in the non-fiction literary award category while its editor-in-chief jointly won the first prize. All these factors added national prestige to the status of the dictionary.

While the editors of the ISN may claim that their dictionary is descriptive, in reality there is no dictionary that can be entirely descriptive. "Dictionaries not only describe a language. Many modern dictionaries are prescriptive works" (Drame 2001: 232). Editors choose which words to define; they also choose how to define and spell them. Even if editors include various pronunciations and spellings for each entry, they still give guidance on which usage is preferable, thus being prescriptive (Hadebe 2006: 29). On the other hand, dictionary users usually assume and expect the dictionary to give them the 'correct' spelling, pronunciation and usage. They furthermore expect the general-purpose dictionary to give them the 'standard' language. Fasold (1984: 247) accentuates this: "Whenever you look up a word in a dictionary to find a correct spelling, you are referring to an authority's language-development decision." What should be noted in this statement is the normative function of dictionaries (Svensén 1993: 45). Chimhundu (1997: 140) says: "A standard dictionary has an important normative influence, as the users accept and apply the lexicographer's descriptions of word-forms and his statements about their meanings. The dictionary is also the chief instrument in interpreting rules about spelling and pronunciation." For a language with little documentation like Ndebele, the production of a monolingual general-purpose dictionary like the ISN contributed to language planning such as, for example, the standardization of spelling, word division and vocabulary. The dictionary tends to be the main reference source to verify the 'correctness' of spelling or to authenticate the usage of some words or expressions (Hadebe 2006: 29).

The current status of the ISN as 'the dictionary' amongst the Ndebele is not peculiar but prevalent in most societies. Generally the monolingual dictionary follows the bilingual dictionary although this is not the case everywhere. For Ndebele, the ISN came after a bilingual English–Ndebele/Ndebele–English publication, *A Practical Ndebele Dictionary* compiled by Pelling. It is interesting to analyze the conditions that led to the rise of the monolingual general-purpose dictionary — often referred to as 'the dictionary'. In Europe, these dictionaries coincided with the introduction of mass education and the development of bourgeoisie nationalism. In Zimbabwe, they became popular after independence, together with nationalist literature in the African languages. The popularity, influence and resultant standardizing function of the ISN should be seen as a consequence of this nationalistic spirit that usually accompanies the emergence of monolingual dictionaries. Similarly, the concerns about language purism discussed below (see 5.1.2) should also be considered from this perspective. Heightened nationalist feelings seem to correlate with heightened language purism.

4. Ndebele in the Zimbabwean Language Landscape

The position of Ndebele within the Zimbabwean language landscape is a precarious one. By *language landscape* is meant "an accurate picture of the languages, both foreign and native, used in a given geographical area" (Ndinga-Koumba-Binza 2005: 133). It should be noted that Zimbabwe does not have a clear language policy (Mkanganwi 1992, Chimhundu 1997). However, English, Ndebele and Shona are recognized in the Constitution as the three main languages. English is the official medium of instruction in Zimbabwean schools although indigenous languages can be used at primary level. Ndebele is only taught as a subject in the two Matabeleland provinces and in some schools in the Midlands province. Shona is taught in the rest of the country.

Zimbabwe also has a number of indigenous African languages designated as 'local languages'², some of which are Tonga, Venda, Nambya, Kalanga and Shangani. Ndebele is in a precarious position because, unlike Shona which enjoys numerical superiority and political prestige, Ndebele is strictly speaking a minority language in Zimbabwe. When compared to Shona, Tonga, Venda or Kalanga, it is a relative newcomer in the history of the country. Almost all the local languages are in Matabeleland where Ndebele is currently dominant (Hadebe 2002: 161). Therefore, a promotion of the status of local languages might be perceived, as previously it has been viewed as an attempt to undermine Ndebele (Hachipola 1998: xx).

Together with the relative advantage of being taught in schools in Matabeleland, the strength of Ndebele is that it serves as the lingua franca for the speakers of local languages (Hadebe 2006: 62). This apparent strength turns out to be a disadvantage for the Ndebele language planner, for example, in the compiling of the ISN, the editors had difficulty in deciding whether or not to

lemmatize loanwords from these other languages. While this seemed the logical way to follow as the dictionary relied heavily on a corpus (see Hadebe 2002), in reality this was not possible. The Ndebele people in general and those of Nguni stock in particular want the Ndebele language to be kept as close as possible to Zulu.

The close links between Zulu and Ndebele seem bound to continue for some time. In any case, Zulu has been and continues to be studied as part of the Ndebele syllabi for secondary education as well as at university level (Hadebe 2006: 62). So the ISN editors had to balance between the scientific evidence provided by the corpus and the sociolinguistic factors dictated by culture, history and speakers' attitudes. The former choice meant including loanwords from local languages and others like Shona, while the latter choice meant keeping the vocabulary very close to Zulu regardless of the fact that no one speaks Zulu.

5. The Ndebele Dictionary, *Isichazamazwi SesiNdebele*

By basing the ISN on a corpus, its compilers aimed to reflect as much as possible the language used by contemporary mother-tongue speakers. In the front matter of the ISN, the editors also reiterate their claim to describing rather than prescribing the language. As already noted above, dictionaries, especially general-purpose standard ones like the ISN, cannot be entirely descriptive. Of course, it has been argued since long ago already that the lexicographer is a historian, not a critic, his duty being to record the inventory of the language without judging it in terms of good or bad (Trench 1857: 3-4). While the editors of the ISN did not necessarily judge some words as good and others as bad, choices had nonetheless to be made resulting in the exclusion of some words for a variety of reasons. According to Follett (1962: 77), "the lexicographer cannot abrogate his authority even if he wants to ... the work itself by its inclusions and exclusions is perceived as the absolute truth". The focus areas here are vocabulary and spelling although these are not the only language planning activities which resulted from the compilation of the ISN.

5.1 Vocabulary Planning

Vikør (1999: 1) says:

To plan the vocabulary of a language is vastly ... complicated, since it consists of a fundamentally infinite mass of units restricted by the limits of our mental capacity. Vocabulary changes and develops spontaneously, more or less, as words change meaning, gradually or more rarely abruptly, or change stylistic value, gradually fall into disuse when they are no longer needed, or are created as neologisms ... This is a development which, one should think, is difficult enough to describe, and virtually impossible to plan or standardize.

At face value, it would seem that Vikør implies that vocabulary cannot be planned, yet he later enumerates a number of ways of vocabulary planning, although difficult to achieve.

The compilers of the ISN did not set out to plan or standardize Ndebele vocabulary, or spelling for that matter, but language planning was the by-product of dictionary-making. In addressing what were lexicographic problems, the compilers found themselves engaged in language planning challenges and had to make decisions that had a profound impact on the standardization of Ndebele.

Some of the aspects of language planning addressed by ISN editors discussed below are norm selection and language purism. The discussion also covers briefly how spelling problems regarding loanwords were dealt with.

5.1.1 Norm Selection

Norm selection is an important aspect of standardization, which involves the choice of one or more varieties to serve as standard reference for a language. For Ndebele to be reflected in the ISN, the problem of norm selection was of a double kind. On the one hand, applied the norm of the Ndebele corpus where choices among the varieties of Zulu had to be made. On the other hand, applied the norm of Zulu where a policy of close relationship to Zulu had to be followed.

Prior to the building of the Ndebele language corpus by the ALLEX Project the general view in Zimbabwe was that Ndebele had no variations. As Chimhundu (1993: 58) put it:

Unlike Shona, Ndebele has no dialects or regional varieties as such. However, there are certain forms and usages that are peculiar to people of particular areas as the result of influences of other languages spoken in those areas, notably Kalanga in Plumtree, Lilima (Tswana?) in Gwanda, Lozi in Hwange, Shangwe in Gokwe-Nkai and Shona in Gweru — Midlands.

This observation is valid only in so far as the variation in Ndebele is not the same as that of Shona and its distinct geographical dialects. In any case, no language can in all aspects be compared with another. Evidence from the Ndebele language corpus shows that there is considerable variation in Ndebele although it is not easy to delineate this variation according to geographical distribution.

As already noted, this apparent confusion can be attributed to the socio-linguistic situation in Matabeleland where Ndebele is spoken alongside other indigenous languages. It could be argued that, although variation in Ndebele is largely attributed to the influence of other indigenous languages and seen as confined to speakers of Ndebele as their second language, there is variation in Ndebele among mother-tongue speakers too. As vocabulary items can come from all these potential sources, it is now the task of the lexicographer to decide whether certain words have been accepted into the mainstream or not.

Although few present-day Ndebele people can trace their ancestry to Zululand, nevertheless almost everyone believes that Ndebele originated from the Zulu language. Because of this, the reasoning is that Ndebele must be kept as close as possible, if not similar to Zulu. The situation is compounded by the fact that Ndebele in Zimbabwe is taught alongside Zulu from high school to university. The teaching of Ndebele, be it grammar, literature, poetry or culture, relies on Zulu books (Hadebe 2006: 136). For this reason, the vocabulary used in formal education and in books is either Zulu or akin to Zulu. Any changes that cannot be accounted for in Zulu are likely to be resisted by a large section of Ndebele speakers (Hadebe 2006: 136). The compilers of the ISN could scarcely ignore the Zulu factor and still hope to produce an acceptable Ndebele dictionary.

5.1.2 Language Purism

Language purism, "a prescriptive approach to language which favours native over foreign words" (Hartmann and James 1998: 113), is one other form of language planning pertaining to vocabulary. In Ndebele, it can be explained in terms of the sociolinguistic factors affecting Ndebele in relation to other languages as well as its historical ties with Zulu (see Ncube 2005: 295). Hadebe (2006: 136) explains this as follows:

As a language, Ndebele found itself surrounded by languages very different from and yet co-existing with it. That may partly explain the high degree of purism as a form of protection from the pervasive influence of the languages in the region. Where the speakers felt threatened by new words from languages in the region, they would instead adopt Zulu words rather than use words from their neighbouring languages. This purism extends even to words from English, which is perceived as superior to Ndebele.

Like all living languages, Ndebele is obviously losing this battle for purism as more and more words are introduced into the language. What is of importance here is how the editors of the ISN dealt with purism. There were two principles that guided selection of entries: first was the frequency of a word as reflected in the concordances drawn from the corpus, and second was the preference for indigenous words over loanwords. From the outset, it was clear that the second principle curtails and overrides the first principle. In practice, it is evident from the Ndebele language corpus that lemmatization was influenced more by what compilers believed were indigenous words than by what the evidence of high-frequency use from the corpus was.

The following are examples of words with a high frequency in the Ndebele corpus that were excluded from the ISN:

ubutindindi (type of mushroom)
izitatangwa (swamps)
-tshotshola (pounding maize)

-*petsha* (to sieve)
 -*kwaya* (to dig)

While these are now common words in spoken Ndebele as reflected in the corpus, they are, however, still considered unacceptable by the more conservative members of Ndebele society, the reason being that since these words derive from Kalanga or its related dialects, they should therefore not be part of Ndebele vocabulary.

There was also another guiding principle in the style manual explicitly stating that in the event of a loanword being synonymous with an indigenous word in the language, the indigenous word should carry the definition. The following are some examples:

indigenous word	loanword
<i>ubabhemi</i>	<i>idonki</i> (adapted from English) (donkey)
<i>umongikazi</i>	<i>unesi</i> (adapted from English) (nurse)
<i>umkhongi</i>	<i>idombo</i> (from Kalanga) (marriage-go-between)

It may be asked: Did the editors apply the principle of higher frequency in the corpus or did they give priority to less used words merely because they sound more original to the language? The answer is clear from evidence in the ISN that the editors gave preference to what was considered a more indigenous word, even if corpus evidence showed overwhelmingly that the word was less used, as was the case with *umkhongi* (compared to *idombo*) under the examples given above. The concern for revivalism meant that words perceived to be indigenous to Ndebele had to be promoted even if that violated the rule of high frequency.

The preference for indigenous words was not pursued to the absolute exclusion of loanwords. Ndebele has borrowed extensively from English, especially words relating to modern fields ranging from agriculture, industry and commerce to sports and religion. Goods that are either factory made or imported come to the Ndebele community with English labels. Although there are attempts to coin indigenous Ndebele words, the process cannot cope with, let alone rival, the loanwords from English. The ISN lists many loanwords from English but, where there is an indigenous Ndebele word equivalent, this carries the definition.

It should be noted that there are some words in everyday spoken Ndebele that are resisted in writing because they are considered not to be Ndebele. One much debated word is the verb *-kwanisa* (be able). Language purists and teachers would insist that the word *-enelisa* (be able) should be used instead as this is considered good and appropriate language. The ISN did not lemmatise *-kwanisa* even though it is more common than *-enelisa*, for fear of it being labelled foreign or 'bad' language. There are also words with sounds that are resisted by some sections of the Ndebele-speaking community, especially the elderly. Although the affricate /dz/ is not original to Ndebele, a number of words with this sound currently exist. Instead of articulating the sound /dz/ some speak-

ers replace it with /j/. Most words with these sounds have become variants like in the following examples:

udzidziyane — *ujjiyane* (blue waxbill)
ubudzugwe — *ubujugwe* (type of mushroom)
-dzimila — *jimila* (to be lost and have memory lapses)

These examples illustrate the problems encountered by editors in headword selection and in defining. When the ISN was published, there was much response to this vocabulary selection. A notable Ndebele writer and historian congratulated the ISN editors for including words with ⟨r⟩:

iringi (ring)
ireshini (ration)
irediyetha (radiator)

Some were of the opposite view, expressing deep concern over the inclusion of words with ⟨r⟩ that they felt were not 'proper' Ndebele. The word they complained most about is *ishamari* (illicit lover), derived from the Shona *shamwari* (friend). From some of the responses, the conclusion can be drawn that Ndebele purism is selective and targeted against words from specific languages whose speakers are disliked or disdained for whatever reason. According to Ncube (2005: 299), "some users of ISN are comfortable to accept the /r/ sound in words such as *irula* (ruler) and *irobhothi* (robot), but are not comfortable when the same sound appears in words such as *ishamari*". It is an observed fact that "vocabulary selection policy inevitably reflects the interests, concerns and culture of the time" (Jackson and Amvela 2000: 163).

What all the above observations show is the overriding considerations for language planning. Since general-purpose standard dictionaries are assumed to be descriptive and not prescriptive, the ISN should then be expected to have lemmatized all high frequency words from the corpus but it did not always do so. Maybe attempts to forestall criticism and rejection from purists led editors to violate the principle of lemmatizing high frequency words from the corpus, instead promoting those words that are generally perceived to be more indigenous and 'authentic' Ndebele. The ISN, like any dictionary of its type, does not only reflect the prevailing language policy, but is itself an agent of language planning, especially with regard to vocabulary.

5.2 Spelling

A big challenge to the editors of the ISN was the spelling of loanwords. In cases where there could be no consensus on spelling, the editor-in-chief had to use his discretion. Loanwords firstly bring into Ndebele sounds that previously were not in the language, making it difficult to represent them in writing (Hadebe 2006: 138). Secondly, loanwords violate the sequences of phonemes acceptable to Ndebele.

The spelling issues discussed here are in two ways important for orthography standardization. Firstly, they show the problems the editors of the ISN had to face. Secondly, whatever decision the editors made has significance for language planning. Unlike the study of vocabulary standardization which remains vague and uncertain, the standardization of spelling is more concrete and direct, thus likely to be easily noticed by language users. However, the complicating factor was that the editors of the ISN had no mandate to make spelling reforms. They did not want to have their dictionary rejected by the authorities on the grounds that it had an unacceptable spelling. It was considered easier to defend the inclusion of particular words in the dictionary against criticism than the violation of certain spelling conventions. The following were some of the problem cases:

(a) The voiced alveolar trill [r]

Originally, the Ndebele phonological system did not have the voiced alveolar trill [r] represented in the Ndebele orthography by the letter ⟨r⟩. The /r/ sound which is not original to Ndebele is now part of it, at least of spoken Ndebele, due to the influence of other languages, especially English and other African languages like Sotho and Shona. Language purists among the Ndebele people are opposed to the use of /r/, and whenever this sound is to be used, they replace it with /l/ written ⟨l⟩.

The controversy about /r/ did not end with arguments on the acceptability of the phoneme as part of the inventory of the Ndebele phonological system but extended to the alphabet. The use of the letter ⟨r⟩ is still being resisted in Ndebele although it seems the /r/ sound is now prevalent in speech. For some reason, it has been assumed that the /r/ sound is a distinguishing feature of Shona while /l/ has been seen as typically Ndebele. It would seem that this perception is mutual between Shona and Ndebele speakers.³ The response to the ⟨r⟩ has already been mentioned above under 5.1.2, together with a discussion of examples.

(b) The voiced alveolar affricate [dz]

Another sound that is still controversial in Ndebele is the voiced alveolar affricate [dz], which is also found in other Zimbabwean languages like Shona, Venda and Kalanga. In some Ndebele words, this ⟨dz⟩ has been substituted by ⟨j⟩. Why it is resisted is unknown. It would seem that the /dz/ has replaced other sounds in Ndebele, as can be seen from the following examples:

<i>ugcigciyane</i> (blue waxbill)	⟨gci⟩
<i>udzidziyane</i>	⟨dz⟩
<i>ujjiyane</i>	⟨j⟩

The original name of the bird is *ugcigciyane* but some speakers probably had difficulties with the click sounds and substituted those with /dz/, saying *udzidziyane*. Now some speakers avoid /dz/, replacing it with /j/ and saying *ujijiyane*.

There are a number of loanwords that have /dz/ substituted by some speakers with /j/, as in the following examples:

<i>ubudzugwe</i> (type of mushroom)	⟨dz⟩
<i>ubujugwe</i>	⟨j⟩
<i>ukudzimila</i> (to be confused, losing one's sense of direction)	⟨dz⟩
<i>ukujimila</i>	⟨j⟩

These words have become either synonymous or are freely variable, both spellings ⟨dz⟩ and ⟨j⟩ having been used in Ndebele.

While some Ndebele speakers resist the affricate /dz/ on the grounds that it is not Nguni and therefore not Ndebele, it is actually found in related Swati although in different phonological environments than in Ndebele. For example, in Swati the voiced alveolar stop /d/ would change into the affricate /dz/ when followed by the vowels /a/, /e/, or /i/.

(c) The voiced postalveolar fricative [ʒ]

The voiced postalveolar fricative [ʒ], orthographically represented in Ndebele by the digraph ⟨zh⟩, which is found in place names like eBezha, eBhazha and eZhilo in Ndebele-speaking areas, has been accepted in Ndebele. Some might argue that these names are in fact not Ndebele but come from the dialects Kalanga or Nyubi. Words spelt with ⟨zh⟩ like *ibhizha* (butter from cowpeas) have been lemmatized in the ISN without any complaints being received.

(d) Unacceptable clusters

The editors of the ISN had to deal with thousands of loanwords containing unacceptable clusters but having become part of everyday language. Because there are currently no comprehensive rules on Ndebele orthography, writers either use italics or the English spelling for these words. In the dictionary, none of these strategies were suitable; the loanwords had to be spelt one way or the other. Below are some examples:

<i>isiliva</i> (silver)
<i>isikweya</i> (square)
<i>-fulasha</i> (flash)

Now that the ISN has been endorsed officially by the Ministry of Education and Culture, which therefore means that it may be used in all schools as a reference book, the spelling and word division employed in the dictionary has become official.

6. Implications for Language Planning in Ndebele

Since the spelling reforms of the 1950s, the production of the ISN was probably the most significant language planning activity in the Ndebele language. In terms of vocabulary development and spelling regulation, the ISN is as yet unparalleled. So far the ISN has undoubtedly been instrumental in the standardization of the Ndebele language, especially as far as vocabulary and spelling are concerned.

What might have given rise to the proliferation of different spellings, especially of loanwords, was the lack of an up-to-date guide to spelling and word division. This was further worsened by the lack of a comprehensive monolingual Ndebele dictionary, which could be used as a reference source for problematic spelling and word division. Although there is still no up-to-date guide to Ndebele orthography, there is a comprehensive Ndebele dictionary (Hadebe 2006: 189) which can at present be employed for reference purposes. Teachers and writers would have to rely on it for 'standard' spelling. In this way, the ISN has become the standardising instrument for Ndebele orthography.

The fact that the ISN has been based on the Ndebele language corpus had the capturing of current language usage as a result. However, the basing of the ISN on the Ndebele language corpus meant the inclusion of some words and meanings perceived by some as foreign or slang or even as 'bad' language. There were concerns about the lemmatization of some words considered not Ndebele. Loanwords, especially from English and other indigenous languages, have become part of everyday Ndebele vocabulary, the question for the lexicographer being whether to lemmatize these or not. Those words included in the dictionary have been endorsed as part of Ndebele while those excluded have been rejected. In this way, the ISN has played a crucial role in Ndebele vocabulary planning.

7. Conclusion

The lexicographer, especially the one compiling a monolingual African language dictionary like the ISN, should take into cognizance language planning issues. It was shown how the ISN has been instrumental in vocabulary planning and spelling standardization in Ndebele. While the compilers of the ISN did not set out to directly engage in language planning, nonetheless their lexicographic activities had significant language planning effects on Ndebele. It was also noted that language planning is goal-oriented, aiming to solve perceived language problems. Lexicographers treating African languages therefore face language planning challenges like codification, spelling reform, norm selection, elaboration and general standardization problems they have to resolve. Furthermore, the function of dictionaries as standard reference works strengthen the role of lexicographers as language planners. In this way, language planning is inevitably linked to monolingual lexicography.

Endnotes

1. The writer as the editor-in-chief of *Isichazamazwi SesiNdebele* is drawing from practical experience. These issues have also been discussed in his doctoral thesis entitled *The Standardisation of the Ndebele Language through Dictionary-Making*.
2. In the Education Amendment Bill, 2005, the term 'minority' has been substituted by 'local' languages and these can now be used as medium of instruction for primary education.
3. For example, the monolingual Shona dictionary *Duramazwi reChiShona* avoided entries with /1/ but the later advanced version *Duramazwi Guru reChiShona* uses the /1/.

References

Dictionaries

- Chimhundu, H. (Ed.).** 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.).** 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Hadebe, S. (Ed.).** 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Hartmann, R.K.K. and G. James.** 1998. *Dictionary of Lexicography*. London: Routledge.
- Pelling, J.** 1966. *A Practical Ndebele Dictionary*. Harare: Longman Zimbabwe.

Other Literature

- Alberts, M.** 2005. A Few Words from AFRILEX. *Lexikos* 15: xiv-xv.
- Alexander, N.** 1992. South Africa: Harmonising Nguni and Sotho. Crawhall, N. (Ed.). *Democratically Speaking: International Perspectives on Language Planning*: 56-68. Cape Town: National Language Project.
- Bamgbose, A.** 1991. *Language and the Nation*. Edinburgh: Edinburgh University Press.
- Béjoint, H.** 2000. *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.
- Chimhundu, H.** 1993. The Status of African Languages in Zimbabwe. *Southern African Political and Economic Monthly* 7(1): 57-59.
- Chimhundu, H.** 1997. Language Standardisation without Policy or Planning: Zimbabwe as a Case Study. Røyneland, U. (Ed.). *Language Contact and Language Conflict*: 129-150. Volda: Volda College.
- Drame, A.** 2001. Foreign Words as a Problem in Standardisation/Lxicography: English and Afrikaans Loan-words in isiXhosa. *Lexikos* 11: 231-241.
- Eastman, C.M.** 1992. Sociolinguistics in Africa: Language Planning. Herbert, R.K. (Ed.). 1992. *Language and Society in Africa: The Theory and Practice of Sociolinguistics*: 1-19. Johannesburg: Witwatersrand University Press.
- Fasold, R.** 1984. *The Sociolinguistics of Society*. Oxford: Blackwell.
- Follett, W.** 1962. Sabotage in Springfield: Webster's Third Edition. *The Atlantic* 209: 73-77.
- Hachipola, S.J.** 1998. *A Survey of the Minority Languages of Zimbabwe*. Harare: University of Zimbabwe Publications.
- Hadebe, S.** 2002. The Ndebele Language Corpus: A Review of Some Factors Influencing the Content of the Corpus. *Lexikos* 12: 159-170.

- Hadebe, S.** 2006. *The Standardisation of the Ndebele Language through Dictionary-Making*. Harare/Oslo: The ALLEX Project.
- Jackson, H. and E.Z. Amvela.** 2000. *Words, Meaning and Vocabulary*. London/New York: Cassell.
- Jernudd, B.H. and J. Das Gupta.** 1971. Towards a Theory of Language Planning. Rubin, J. and B.H. Jernudd (Eds.). *Can Language Be Planned? Sociolinguistic Theory and Practice for Developing Nations*: 195-216. Honolulu: The University Press of Hawaii.
- Mkanganwi, K.** 1992. Language Planning in Southern Africa. Crawhall, N. (Ed.). *Democratically Speaking: International Perspectives on Language Planning*: 6-11. Cape Town: National Language Project.
- Ncube, C.** 2005. Language Development or Language Corruption? The Case of Loan-words in *Isichazamazwi SesiNdebele*. *Lexikos* 15: 294-305.
- Ndinga-Koumba-Binza, H.S.** 2005. Considering a Lexicographic Plan for Gabon within the Gabonese Language Landscape. *Lexikos* 15: 132-150.
- Ntshangase, D.** 1997. Society in Transition. *The Feasibility of Technical Language Development in the African Languages*: 17-20. Pretoria: Department of Arts, Culture, Science and Technology.
- Pauwels, A.** 1998. *Women Changing Language*. London/New York: Longman.
- Rubin, J. and B.H. Jernudd.** 1971. Language Planning as an Element in Modernization. Rubin, J. and B.H. Jernudd (Eds.). 1971. *Can Language Be Planned? Sociolinguistic Theory and Practice for Developing Nations*: xiii-xxiv. Honolulu: The University Press of Hawaii.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Trench, R.C.** 1857. On Some Deficiencies in Our English Dictionaries. *Transactions of the Philological Society* (1857): 1-60
- Vikør, L.S.** 1999. *Vocabulary Planning — A Tricky Business*. Lecture at the ALRI Launching Seminar, Harare.
- Wolff, H.E.** 2000. Language and Society. Heine, B. and D. Nurse (Eds.). 2000. *African Languages: An Introduction*: 298-347. Cambridge: Cambridge University Press.

Considerations for Providing Etymological Information in the *Advanced Ndebele Dictionary**

Langa Khumalo, *African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe (la_nga@yahoo.co.uk)*

Abstract: This article discusses the presentation of etymological information in the prospective *Advanced Ndebele Dictionary* (henceforth the AND) which is still in its planning stage. The AND will be a successor to *Isichazamazwi SesiNdebele* (henceforth the ISN), a medium-sized, general-purpose, monolingual Ndebele dictionary published in 2001. The AND will not just be larger than the ISN, but will be more advanced with regard to the depth and scope of its lexical items and definitions. Whereas the ISN was based on a corpus size of about a million running words, the AND is envisaged to have as basis a corpus size of about five million. To this end, corpus expansion is an ongoing process. Again, unlike its forerunner, the AND will provide additional grammatical information for each lexical item, including phonetic transcription, tone marking and etymology. It is the inclusion of etymological information in the AND that is critically examined in this article. There is a long tradition of providing etymological information in modern dictionaries from Oxford's *New English Dictionary* to ALRI's *Duramazwi Guru reChiShona*.

Keywords: ETYMOLOGY, ADVANCED NDEBELE DICTIONARY, CULTURE, HERITAGE, LANGUAGE CONTACT

Opsomming: Oorwegings vir die verskaffing van etimologiese inligting in die *Gevorderde Ndebelewoordeboek*. Hierdie artikel bespreek die aanbieding van etimologiese inligting in die toekomstige *Gevorderde Ndebelewoordeboek* (voortaan die GNW) wat nog in sy beplanningstadium is. Die GNW sal 'n opvolger wees van die *Isichazamazwi SesiNdebele* (voortaan die ISN), 'n middelgroot, meerdoelige, eentalige Ndebelewoordeboek wat in 2001 gepubliseer is. Die GNW sal nie net groter as die ISN wees nie, maar sal meer gevorderd wees wat betref die diepte en omvang van sy leksikale items en definisies. Waar die ISN gebaseer was op 'n korpusgrootte van ongeveer 'n miljoen woordoptekeninge, word voorsien dat die GNW 'n korpusgrootte van ongeveer vyf miljoen as basis sal hê. Vir hierdie doel is korpusuitbreiding 'n voortgaande proses. Weer eens, anders as sy voorloper, sal die GNW bykomende grammatiese inligting oor elke leksikale item verskaf, insluitende fonetiese transkripsie, toonaanduiding en etimologie. Dit is die insluiting van etimologiese inligting in die GNW wat in hierdie artikel krities beskou word. Daar is 'n lang tradisie vir die verskaffing van etimologiese inligting in moderne woordeboeke vanaf Oxford se *New English Dictionary* tot ALRI se *Duramazwi Guru reChiShona*.

* This article was presented as a paper at the Tenth International Conference of the African Association for Lexicography, organised by the Sesiu sa Sesotho Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

Sleutelwoorde: ETIMOLOGIE, GEVORDERDE NDEBELEWOORDEBOEK, KULTUUR, ERFENIS, TAALKONTAK

Introduction

This article consists of two main parts. In the first part, it will be argued that, except for its interest to the target users of the prospective *Advanced Ndebele Dictionary* (henceforth the AND), etymological information is often valuable as an accompaniment to historical and cultural research. Ndebele shares its heritage with other Nguni languages spoken in South Africa, of which all drew much from contact with English and Afrikaans. Etymological information is valuable, especially for a language like Ndebele that has not been extensively documented. In the second part, challenges with regard to etymological information will be discussed. Historical changes in meaning are an inevitable fact, but are sometimes decried as if language ought to be arrested at some period in time. Consequentially an attempt to fix meanings or to tie words to their "original" meanings is denying the social reality of linguistic usage. A classic example of the lawyer who disputed a witness's use of the word *hysterical*, is told by Warburg (1968: 351-352).

Furthermore, it will be argued that older senses of some words sometimes survive in limited contexts, while the most frequent meaning has changed. As a result, this presents the problem whether to provide such etymological information. Then occasionally there are also erroneous origins that have become enshrined in the language through the process of "folk etymology" by which the pronunciation or spelling of a word is modified on a false analogy. Although it hardly needs to be argued that modern lexicographers try to avoid speculation and guesswork, only giving information based on thorough research, instances such as these have occurred in Ndebele. Finally it will be argued that the historical information can be interesting and instructive, although not all of it is accessible and relevant. The history of a word is not necessarily its meaning.

Background

The background of the Ndebele people has recently been related in a number of works, which include Hadebe et al. (2001), Hadebe (2002), and Khumalo (2003). What follows is a brief synopsis. Around 1820, the Ndebele people broke away from the then mighty Zulu kingdom, in the present-day KwaZulu-Natal province of South Africa. The original group that left Zululand was initially called the Khumalos as their leader, Mzilikazi, was from the Khumalo clan. Mzilikazi and his Khumalos, as they were known then, moved northwards from Zululand into Sotho territory in the present-day Gauteng province of South Africa. Mzilikazi assimilated a number of Sotho people who soon out-

numbered the original Khumalos who were of Nguni origin. Consequently the group ceased to be referred to as the Khumalos. The Sotho called all raiding Nguni groups *kiMatebele*, which is thought to be the origin of the name *Ndebele*. Around 1838–1840, the Ndebele people settled in the southern part of present-day Zimbabwe. Mzilikazi continued to expand his nation by raiding many tribes, incorporating them into his political system. It is in Zimbabwe that the Ndebele people assimilated by far the largest number of people. These mainly included the Nyubi, Kalanga, Nanzwa, Nambya, to some extent the Tonga and a number of Shona tribes. Most of these were previously under rulers known as the Mambo. In 1896, the Ndebele people came under British rule, exposed to yet another language, English, and the technological advancement it brought with it. This scenario presented an environment for language contact and inevitably language change (Khumalo 2003: 2). The challenge in tracing etymological information would be to correctly document this expansive language contact scenario.

Value and Role of Etymology

Landau (2001: 127) writes that people mostly find dictionaries fascinating, particularly the derivation of words presented in them. Etymology is therefore viewed as attracting attention in a dictionary, providing interesting background information about lexical items. From the literature, it is clear that, although dictionary users most often search for meaning, meaning does not have the same distinction as etymology, which combines knowledge of other languages with scholarship, which reveals hidden relationships.

An example to illustrate this observation is the Ndebele word *uMantshingelane* (a guard). The etymology of this word is the English verb phrase *marching on the line*, which is one of the drills the guards perform during their inspection. The word was adapted by breaking up the English sequence, unusual in Ndebele, to form the normal Ndebele sequence, a process that can be presented as follows:

word/phrase	→	>	new word
(from language X)	(decomposition)	(reconstruction)	(in language Y) ¹

The decomposition and reconstruction of the word so that it conforms to the phonological rules of the borrowing language can loosely be referred to as *phonologizing* the new word, the process by which *uMantshingelane* was adapted from English as illustrated below:

English verb phrase	→	>	Ndebele noun
<i>marching on the line</i>			<i>uMantshingelane</i>

It is also clear from this example that the chief value of etymology lies not only in its historical, but also in its linguistic interest. It suggests curiosity about, if

not familiarity with, the study of classical languages, proto-languages or elements of comparative philology (Landau 2001: 131-132).

In the literature, there have been arguments about the usefulness of etymologies. It has been suggested that etymologies should be omitted altogether from non-historical general dictionaries. What then is the role of etymology?

Firstly: Etymological information can provide raw material for the scholar and the student studying the history of a language. To this end, data are provided that lead to further inquiry into a people's history. For example, the etymological information on the word *Ndebele* provides such material. As has been stated above, Mzilikazi had contact with the Sotho with the result that the group ceased to be referred to as the Khumalos. The Sotho referred to all raiding Nguni groups as *kiMatebele* (meaning 'warriors with long shields'), hence the name *Matebele* for the Ndebele (Hughes and Van Velsen 1954: 42), which in Nguni became *amaNdebele*. Nyathi (1994: 24) on the other hand claims that the name *Matebele* derives from the Sotho verb *hotebela*, "okutsho ukubhidliza, ukubulala kumbe ukuchitha", meaning 'to destroy, kill or ruin'. In early missionary writings, the Ndebele are often referred to as the Matabele. It seems the term *Matabele* had been applied to another Nguni group, the Hlubi, long before it was used to designate the present-day Zimbabwean Ndebele people (Dube 1993: 3). It can therefore be concluded that the actual meaning of *Matebele* remains unresolved.

If it is accepted that the name *Ndebele* was derived from *Matebele*, in historical and linguistic studies on Ndebele, it is still a controversy how it happened. It should be noted that in present-day South Africa there is another Nguni group also called *Ndebele*. It is believed that this group left Zululand much earlier and was given the name long before Mzilikazi's group. Rasmussen (1977: 162) writes:

Eventually the name 'Matabele', or 'Ndebele' in its Anglo/Nguni form, came to apply only to Mzilikazi's people and to the 'Transvaal Ndebele'. These latter were the descendants of much earlier Nguni immigrants onto the highveld. Mzilikazi had dealings with some of these communities during the 1820s; however, his people and the Transvaal Ndebele are essentially unrelated.

This information provides material for someone studying the history of the language.

Secondly: Etymological information increases the understanding of and simultaneously stimulates interest both in language in general and in one language in particular. While certain words may be taken for granted because they mean what society has come to accept, etymological information brings an interesting account of how that meaning was arrived at. For example, the etymological information of the word *umfundisi* (reverend and/or teacher) provides such an exposition. Education was introduced to the Ndebele by the missionaries who preached as well as taught in their newly-established schools. The preacher, who taught the scriptural lessons and church hymns to the Nde-

bele converts was also a "linguist", who developed orthographies for the then unwritten language, hence he was seen as a school teacher as well. The word *umfundisi* is derived from the verb *-fundisa* (to teach). When the colonial government in Rhodesia began to build schools for Africans in the 1960s, teachers who were not necessarily preachers, were hired. Today, there is no longer an obvious link between education and the church in Zimbabwe, teachers being seen as independent of the clergy. The word *umfundisi* has now narrowed its meaning, referring to a preacher only. At present a teacher is called *umbalisi*, derived from *-balisa* 'to help to read' (Hadebe 2002: 126). Interestingly though, 'The boy is reading a book' is rendered in Zulu as *Umfana ufunda incwadi*, distinct from Ndebele *Umfana ubala ugwalo*. Having narrowed the sense of *umfundisi*, Ndebele uses the verb *-funda* with the meaning 'to learn'.

Thirdly: Since a dictionary is a record of the culture of the speakers of the language it describes, etymological information also provides clues to the history of a specific culture and its relationship to other cultures. Etymological information of words like *uMlimu* (God) and Ndebele surnames *Mdlongwa*, *Dawu* and *Mnkandla* provide evidence of the historical relationship between the Ndebeles (then known as the Khumalos) and the Sotho. As has been stated above, when the Khumalo clan left Zululand, it had contact with other tribal groups. Between 1822 and 1837, the Ndebeles sojourned in the present-day Gauteng province in South Africa. One event associated with the stay in the then Transvaal region was contact with the missionaries. The London Missionary Society had established a mission station at Kuruman among the Tlaping. Although the Ndebeles did not generally accept the Christian faith, some Sotho words from the religious realm found their way into the Ndebele language, for example, the Sotho name *Modimo* in its Ndebele adoptive form *uMlimu*², the equivalent of English *God*. Although *uMlimu* does not semantically capture God's greatness in the same way as *uNkulunkulu* (literally *the Great Great One*), it has found its place in the Ndebele lexicon, its etymology being understood within this cross-cultural context. Furthermore, the Sotho people, who were incorporated into the fast-growing Ndebele empire abandoned their language in preference to Ndebele. Even their tribal names were changed in conformance with the phonological structure of the Ndebele language, for instance, the Sotho names *Motlokwa*, *Tau* and *Mokgatla* became *Mdlongwa*, *Dawu* and *Mnkandla*³ respectively. It is clear that change, which was inevitable, favoured the dominating tribe (Khumalo 2004: 4).

The Ndebeles have always viewed themselves in many ways similar to the Zulus, arguments sometimes being put forward that these groups belong to one and the same people. According to Hadebe (2002: 129), there are many words assumed to be derived from Zulu, yet actually Zulu adopted them from Afrikaans, thereby becoming part of the Ndebele vocabulary. Lexical etymology of some of the words include *isitulo* (chair) from Afrikaans *stoel*, *ibhatshi* (jacket) from Afrikaans *baadjie*, *ijazi* (coat) from Afrikaans *jas*, *umfolo* (fallow) from Afrikaans *voor* and *umjaho* (chase) from Afrikaans *jaag*. Hadebe (2002: 129)

observes that in the 1970s, there was a period when loanwords from English were avoided in Ndebele. This period of purism saw the substitution of loanwords from English by words believed to be indigenous as they were taken from Zulu to which most Ndebele speakers look as the source of indigenous vocabulary. It is unfortunate that this purism did not purge Ndebele of European loanwords as intended, but substituted English for Afrikaans as the source language. This does not mean that Ndebele speakers preferred Afrikaans to English, but that they adopted these Afrikaans words via Zulu believing them to be indigenous Zulu words. The cause for these Afrikaans adoptions was the contact of the Zulu people with the Afrikaners. More importantly, the etymology of these words is based on a speculative and erroneous perception. This also causes semantic problems to arise. For instance, the meaning of *umjaho* is 'competition' in Zulu and Ndebele, but etymologically the word *jaag* means 'chase' in Afrikaans from which it is derived.

There are also examples which demonstrate cultural contact. One of the greatest changes to the Ndebele language was brought about after the arrival of the Ndebeles in Zimbabwe. As stated above, this was the time when they came into contact with the Nyubi, Kalanga, Nanzwa, Nambya, to some extent the Tonga and a number of Shona tribes. The honorific plural or pronouns of power are one example of Shona lingual influence on the Ndebele language. Among the Shona people, when an older person or a person in a senior social position is addressed, an honorific plural is used. The Ndebele language has embraced this concept by adding the plural morpheme /-li/ as demonstrated by the following two examples:

- (1) *sakubona?* how do you do? (singular)
**salibonani?* how do you do? (plural) [derived from Shona *mhoroi*, which denotes plural]
- (2) *kunjani?* how are you? (singular)
**linjani?* how are you? (plural) [derived from Shona *makadii*, which denotes plural]

Another form of change that has occurred in the Ndebele language is in its verbal morphology as can be seen from the etymology of the words *umile* (he/she is standing) and *hambisa* (hurry up). The etymology of these two words can also be explained by historical cultural contact. The form *umi* (he/she is standing) has been changed to **umile* (derived from Shona *amire*) by adding the extension *-ile*⁴ to the verb stem *umi*. The verb *tshitsha* (hurry) now has an equivalent **hambisa* (derived from Shona *fambisa*).⁵ The Shona word for *hurry up* is *fambisa*, derived from the verb *famba* (walk) by adding an extension *-isa*. The Ndebele language has derived a new word, meaning 'hurry up', from the verb *hamba* (walk) by adding an extension *-isa* to derive the verb *hambisa*.

The conclusion can be drawn from the above that to include entries without etymology in a dictionary is to divorce the language from its past. Further-

more, by doing this, the opportunity is missed to show language in its context, notwithstanding the fact that words without etymology may be very well defined. A dictionary should give users the opportunity to expand their view of the language by offering more information than is necessary, even though some users might have difficulty understanding this. Etymologies demonstrate that language changes in form and meaning. They remind users of the way other cultures, or their own, viewed particular words in the past. Hence, etymology forms an important component of a dictionary.

Problems and Challenges of Providing Etymological Information

While etymological information may be viewed as valuable in its own right, it is often argued that it tells users little about the current meaning of words and is in fact often misleading. The error of associating present meaning with past etymological meaning is commonplace, also in the sciences, particularly medicine (Landau 2001: 128). Manuila (1981), quoted by Landau, points out: "Many words, whether technical or in ordinary use, have departed so widely from their original meaning that the latter can be of interest only to medical historians and linguists."

Another argument against the inclusion of etymological information is that users interested in etymology regard etymologies as stories about word histories. Furthermore, and more importantly, the conventional dictionary treatment of etymologies usually does not shed much light on the semantic development of words, for two reasons:

- (a) As there has often been too little scholarly research, there is uncertainty or ignorance about many etymologies.
- (b) Because dictionaries have better uses for the space, their etymologies have usually been compressed as much as possible.

One challenge facing a lexicographer when planning to include etymological information in a dictionary like the AND, is first to look at the type of product aimed at, and then to consider the primary target users of this product. The AND is a successor to *Isichazamazwi SesiNdebele* (henceforth the ISN), a monolingual dictionary with 20 010 headwords targeted mainly at secondary school users. The primary target users of the AND will be mainly tertiary institutions, including high schools, colleges and universities. These form a group whose research is more investigative and more profound in disciplines such as language and history. It should be considered whether it would have positive value to give etymological information in a dictionary targeted at this group. The editors have decided to include etymological information, primarily because they recognise that a dictionary is a cultural repository in which information on the origin of words is an accompaniment to linguistic and historical research.

Planning is one challenge and implementing or executing the plan is another. A dictionary has to balance both detail and space. To this end, one of the challenges is the extent to which the editors have space to include the etymological detail of each word. The information should be detailed enough to give full scope to the value of etymology, but at the same time remain balanced with respect to space. The question the editors should constantly consider is how much etymological detail should be included in the AND.

Another challenge pertains to the documentation of the language. Ndebele has not been extensively recorded. It is only recently that some works on Ndebele have been published, including the ISN, its first monolingual dictionary which appeared in 2001. The challenge therefore is how much historical facts and details can possibly be found for a language like Ndebele which is not well documented. This is in the light of the existing scepticism about oral tradition which is steeped in myths.

The solution to this challenge is that enough time should be allowed the editors for research to verify facts and details. Planning should be well in advance to give the editors enough time for doing research. Consequently, including etymological information in the AND will certainly stimulate research in the language.

Conclusion

That a dictionary is an important resource for the preservation of a people's cultural heritage for posterity is indisputably true. In adding value to this function, the tradition of including etymological information in modern dictionaries has developed for a considerable time already. A dictionary should not just be a list of words with their definitions exemplified in context, but it should be more. An interest in knowing the origin of words has been established. This has added a new dimension to dictionaries, increasing their research value and in the process expanding their scope and uses. This article has tried to demonstrate the importance of giving etymological information in the AND, in spite of certain challenges and problems.

Endnotes

1. X indicates the language from which Y is borrowing.
2. The Ndebele language has a noun class prefix system. The prefix morpheme *u-* is added to the adopted word.
3. These are now well-known Ndebele surnames. In conformity with the Ndebele phonological system, the Sotho consonant changes from /t/ to /d/, /k/ to /ng/ and /kg/ to /nk/.
4. Ndebele orthography does not allow the consonant /r/, so in all adopted words the letter *r* changes to *l*.
5. Note that * represents a construction influenced by Shona. Although such constructions are still resisted in certain quarters, they are very prevalent.

References

- Chimhundu, H. (Ed.).** 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Dube, C.** 1993. Mzilikazi's History — The Logic. Letter. *Moto Magazine* 124: 3, May.
- Hadebe, S.** 2002. *The Standardisation of the Ndebele Language through Dictionary-Making*. D.Phil. Thesis. Harare/Oslo: University of Zimbabwe/University of Oslo.
- Hadebe, S. (Ed.).** 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Hughes, A.J.B and J. van Velsen.** 1954. *The Shona and the Ndebele of Southern Rhodesia*. London: International African Institute.
- Khumalo, L.** 2003. *A General Introduction to Ndebele Grammar*. Cape Town: CASAS.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. Cambridge: Cambridge University Press.
- Manuila, A. (Ed.).** 1981. *Progress in Medical Terminology*. Basle: S. Karger.
- Nyathi, P.** 1994. *Igugu LikaMthwakazi: Imbali YamaNdebele 1820–1893*. Gweru: Mambo Press.
- Rasmussen, R.K.** 1977. *Mzilikazi of the Ndebele*. London: Heinemann.
- Warburg, I.** 1968. Notions of Correctness. Quirk, R. 1968². *The Use of English*: 347-359. London/Harlow: Longman.

Lexicographical Practice and Lexicological Research: The Case of Shangani in Zimbabwe*

Peniah Mabaso, *African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe* (penmabaso@arts.uz.ac.zw)

Abstract: The article outlines the contribution of ALRI towards the development of the minority languages in Zimbabwe, with special reference to Shangani. After a description of the status of Shangani, a brief summary of the Zimbabwean language use policy is given. The challenges for the development of lexicographical practice and lexicological research are discussed with regard to a minority language that has been marginalized for a long time. After the consideration of the kind of dictionary planned for Shangani in accordance with the user profile, the role of a dictionary in codifying and harmonizing a language is stressed. Except for contributing to the orthographic development of a language, lexicography also forms the basis for the furtherance of literacy in a community.

Keywords: ALRI, MINORITY LANGUAGES, MARGINALIZED LANGUAGES, LEXICOGRAPHICAL PRACTICE, LEXICOLOGICAL RESEARCH, DICTIONARY, TARGET USERS, SHANGANI

Opsomming: Leksikografiese praktyk en leksikologiese navorsing: Die geval van Sjangani in Zimbabwe. Die artikel skets die bydrae van ALRI tot die ontwikkeling van die minderheidstale in Zimbabwe, met spesiale verwysing na Sjangani. Na 'n beskrywing van die status van Sjangani, word 'n kort opsomming van die Zimbabwiese taalgebruiksbeleid gegee. Die uitdagings vir die ontwikkeling van leksikografiese beleid en leksikologiese navorsing word bespreek met betrekking tot 'n minderheidstale wat vir 'n lang tyd gemarginaliseer is. Na 'n beskouing van die soort woordeboek wat in ooreenstemming met die gebruikersprofiel vir Sjangani beplan word, word die rol van 'n woordeboek vir die kodifisering en harmonisering van 'n taal beklemtoon. Behalwe dat dit bydra tot die ortografiese ontwikkeling van 'n taal, vorm leksikografie ook die basis vir die bevordering van geletterdheid in 'n gemeenskap.

Sleutelwoorde: ALRI, MINDERHEIDSTALE, GEMARGINALISEERDE TALE, LEKSIKOGRAFIESE PRAKTYK, LEKSIKOLOGIESE NAVORSING, WOORDEBOEK, TEIKENGEBRUIKERS, SJANGANI

* This article is based on 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.

1. Introduction

Shangani, also called Tsonga in South Africa and Changana or Hlengwe in Mozambique, is categorized in the Zimbabwe Education Act 1987 as a minority language because of its number of speakers. This Act is also used as the country's language policy. During the colonial period, Shangani never had a significant role in education. The language was therefore neglected and marginalized until the 1982 consultations on the Zimbabwean languages. These consultations indicated that for most of what became to be recognized as minority languages, neglect and marginalization had had a very negative impact on their existence. The adoption of the majority languages in Zimbabwe as languages of education meant that the speakers of these neglected and marginalized languages were assimilated into the speech communities of the majority languages. The Shanganis in the Shona area came to be regarded as Shona speakers. And similarly, the Kalangas in the Ndebele area came to be considered as Ndebele speakers. So the lack of minority language use meant that many speakers also lost their ethnicity and the knowledge of their languages.

The establishment of the African Languages Research Institute (ALRI) in 2001 provided the impetus for developing the previously neglected and marginalized minority languages. The focus has been on lexicological and lexicographical research as a way of furthering the indigenous languages. Shangani is the latest to embark on this development. The delay in the development of Shangani can be explained by many impeding factors. The most important among these factors are the neglect and marginalization of the language after colonialism and the attendant language use policies in schools. Also there was no researcher at ALRI who could undertake work on the language. The lack of resources is always an important consideration in the development of minority languages.

The aim of this article is to discuss the challenges facing the furthering of lexicological and lexicographical research in minority languages like Shangani. As these challenges are numerous, the article will focus on research at ALRI and the training of lexicographers. Without these developments, the language cannot possibly contend to representation in education. Lexicography as objective at ALRI is one of the viable means to codify and harmonize the language, also to form the basis for other developments such as orthography and literacy. The fact that Shangani is also spoken in South Africa and Mozambique introduces another critical dimension to its linguistic study and the development of its lexicological and lexicographical research. Each of the countries where the language is spoken has its own orthography and agenda for development. As the work proceeds, major choices will have to be made. Not to consider Shangani language developments in countries outside Zimbabwe may end up in repeating and doubling research. Consulting with researchers in Shangani-speaking communities in neighbouring countries may entail generalizations through harmonization strategies, possibly with the loss of the specific indige-

nous content of Zimbabwean Shangani. This situation therefore causes a dilemma for the development of Shangani, which, apart from receiving no resource allocations, also has few researchers working collaboratively on all aspects of the language.

2.1 Who are the Shanganis?

The Shanganis are a minority ethnic group concentrated in south eastern Zimbabwe. They are found in the southern districts of Masvingo Province, mainly Chiredzi South and Mwenezi. There are also very few numbers of them in other districts like Zaka, Masvingo South and Mberengwa. Shangani-speaking communities also reside in Southern African countries like Mozambique, South Africa and Swaziland. In Mozambique, they are referred to as the Changana or Hlengwe, while in South Africa, they are known as the Tsonga. Hachipola (1998) notes that the term *Shangani* is a generic name used to refer to the followers of Soshangana, who was one of Zwide's military leaders. Soshangana, a Tsonga himself, migrated northward and settled in present-day Zimbabwe. On his way to the North, Soshangana incorporated other smaller groups like the Ndau, Ngoni and Nyembeni (Hachipola 1998). As a result, Zimbabwean Shangani has a very close linguistic relationship with Tsonga spoken in neighbouring South Africa and Changana or Hlengwe spoken in Mozambique.

It can be argued that Zimbabwean Shangani is a Tsonga variety caused by historical and linguistic factors. There exists a great degree of mutual intelligibility between Tsonga and Zimbabwean Shangani which suggests a closer relationship than mere linguistic contact. Zimbabwean Shangani now has some peculiar features of phonology and vocabulary which differ from Tsonga. This is understandable as language, which is as dynamic as society itself, changes over a period of time. However, on this basis it cannot be argued that Zimbabwean Shangani should be viewed as an independent variant of Tsonga, and thus has to benefit from a specific linguistic and cultural study. Also, as the three countries where Tsonga is spoken are sovereign states with their own development policies, it is not necessarily easier to undertake mutual beneficial collaborative studies. This would only be possible when some research on Zimbabwean Shangani has been done.

2.2 What is the status of Shangani in the respective regions?

In Zimbabwe, Shangani is one of the country's officially recognized minority languages, spoken by around five percent of the total population. However, with its minority status, it has not benefited much since the colonial era from the Zimbabwean language policies. Although it is a language of significant presence in South Africa and Mozambique, there are no inter-border links that could promote the inter-country usage of Shangani. In South Africa, the lan-

guage is among the country's eleven official languages and in Mozambique, where it is also well developed, it is one of the country's official national languages. On the other hand, because of particular historical circumstances, Shangani is an undeveloped minority language in Zimbabwe. This means it could not benefit from linguistic development allocations by the government. The consequences of this are unfortunate for the language. Shangani is mostly maintained in family and community circles, chiefly as a spoken language with no publications produced in it in Zimbabwe. The only written material in it consists of texts such as informal letters by literate speakers using their own orthographies. Zimbabwean Shangani lags behind in orthography mainly because it has been scantily studied by linguists. The government also paid very little attention to the development of the country's minority languages. Speakers of minority languages are forced to use either Shona or Ndebele, the national languages. As a result, most minority language speakers are bilingual. While bilingualism is an advantage in a situation where language policies do not protect or promote minority languages, they are eventually abandoned with the result that the speakers are ethnically assimilated into the speaker communities of the majority languages.

3. Zimbabwe's current language use policy

Zimbabwe's language use policy played a big role in subverting the development of the country's minority languages. Viriri (2003) notes that according to the wording of the Education Act these minority languages seem not to exist. The Zimbabwe Education Act of 1987, Chapter 55, which doubles as the country's language policy states that:

- (a) Subject to the provisions of this section, the three main languages of Zimbabwe, namely Shona, Ndebele and English shall be taught in all primary schools from the first grade as follows:
 - (i) Shona and English in all areas where the mother tongue of the majority of the residents is Shona or
 - (ii) Ndebele and English in all areas where the mother tongue of the majority residents is Ndebele.
- (b) Prior to fourth grade either of the languages preferred may be used as the medium of instruction depending upon which language is more commonly used and better understood by the pupils.
- (c) From the fourth grade, English shall be the medium of instruction provided that Shona or Ndebele shall be taught as subjects on an equal time allocation basis with the English language.
- (d) In areas where minority languages exist, the Minister may authorize the teaching of such languages in primary schools in addition to those specified in subsection (a), (b) and (c).

From these stipulations, it becomes clear that English has the higher status in Zimbabwe's Education Act. It is the most prestigious language, with the African languages being given the lower status. Chitiga (1995) traces this situation back to the early days of the colonization of Zimbabwe, arguing that because it was the colonizer's language, English assumed the most important position in the life of Zimbabweans in all formal spheres like education, commerce and the media.

Since it is taught in all regions, English is placed at the top of the hierarchy, followed by Shona and Ndebele which are instructed in their respective regions, even up to tertiary levels. Minority languages like Shangani are at the bottom of the hierarchy since they are taught only up to grade three. However, it should be noted that Shona and Ndebele used to be taught from ordinary level up to tertiary level through medium of English. It is only since February 2006 when the Zimbabwe School Examination Council (ZIMSEC) sent a circular that English should no longer be used in teaching Shona and Ndebele that this situation has changed.

Apart from being the only publication languages, Shona and Ndebele are given much airtime on national radio stations, while the other recognized minority languages are broadcast on National FM with news slots of only five minutes. Both Shona and Ndebele are also used on this radio station. On National Television, English is given more airtime than any other language. Airtime allotted to English news bulletins ranges from thirty minutes to over an hour during the course of the day while Shona and Ndebele news bulletins are given fifteen minute slots twice daily. Discussions on current affairs are usually presented in English. There are also drama programmes in Shona and Ndebele on National Television. Minority languages like Shangani do not have these privileges.

This shows that there is much that needs to be done for these minority languages. In the case of Shangani, speakers are unable to study their language beyond primary school level in post-colonial Zimbabwe. Minority languages like Shangani have never been examined at any level. In the 1930s and 1940s, Shangani was taught up to standard six at the Lundi Mission School but learners never sat for examinations. It was only in 2004 that the government revised the Zimbabwe Education Act, making provision for the teaching and examination of Shangani and the other five official minority languages at school and tertiary education level. However, the government does not seem to take significant action in promoting the teaching and examination of these languages. There are very few government projects for the development of these languages. The government does not seem to be taking issues of language development seriously, probably because it is occupied with other issues considered of more importance.

Another major problem is the unavailability of written materials for use in schools. The teachers responsible for these languages have to translate either Shona or Ndebele texts. All the minority languages that are taught have no

standard orthographies. The teachers use diverse orthographies. ALRI has now intervened to improve the Zimbabwean language situation.

4. ALRI'S role in language education and development

ALRI is dedicated to researching, documenting, developing and promoting Zimbabwe's indigenous languages. It also aims at promoting the marginalized minority languages, Shangani included. ALRI's chief preoccupation is the compilation of dictionaries in the indigenous languages. So far it has published dictionaries in the two major indigenous languages, Shona and Ndebele.

Compiling a dictionary does not guarantee its proper usage. Dictionary-using skills must be learnt to appropriately respond to community needs. Users must be made to realize the importance of dictionaries. Also, language use policies facilitating communication between languages need to be in place for dictionaries, especially bilingual ones. If there is no recognition of formal literacy in other languages, there will be general illiteracy in communities speaking the indigenous languages. The following observation from the report of the World Commission on Culture and Development (1995: 57) should be noted:

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

From this observation, it becomes clear that the policies denying the rights of minorities in the areas of culture and education, and exerting domination by linguistic and ethnic hegemonies, create a state of discrimination and denial of access to self-actualization. To break this vicious circle, this report urged the global community to adopt a vision for the 21st century that would uphold the protection and exercise of cultural rights, especially in education. The development of resources to support these is critical for the survival of minority languages. For Zimbabwe this means that minority languages like Shangani must also be researched, documented, developed and promoted. The following sum up the reasons for and the objectives of the ALRI Shangani dictionary project:

- As dictionaries are increasingly becoming effective tools for storing knowledge and its transmission, and improving language and culture acquisition and appreciation, Shangani lexicographical development will be instrumental to a revival of the language.
- Skilful dictionary use will contribute to better interlanguage communication and more independence and confidence in language learning and

culture information, especially in the case of Shangani which is spoken as a minority language.

- How well dictionary skills in language and culture communication can be learned will depend on the way the dictionary presents its material. So it is hoped that the dictionary will be the integral text in education in Shangani, and that it will afford learners to explore and appreciate their language.

What ALRI has actually demonstrated with the publication of the Shona and Ndebele dictionaries is that a dictionary has hitherto been an unknown tool for the speakers of these languages. As a tool for storing knowledge about a language and its culture and for supporting literacy and learning, speakers have come to realize the authoritativeness of a dictionary. Speakers of Shona and Ndebele who could never before have been persuaded that they need such a tool, now realize the value of a dictionary. Hopefully Shangani speakers will also see in their dictionary the benefits of developing their language.

There are several challenges to the use of Shangani dictionaries in the Zimbabwean educational system:

- School syllabi are designed by people who have never had formal training in dictionary use: dictionaries are seen as works of reference, only needed when the problem of a word's meaning arises; dictionaries are thought to be confined to learners, who lack fluency in a language; and dictionaries are never thought of as providing any practical skills.
- Unavailability of dictionaries is never felt to be a very serious problem: the language teacher is thought to fulfil the language needs of the pupils and students; and the dictionary is viewed as a luxury so that its place is difficult to define in the teaching-learning process.
- Poor reference skills among teachers and their scholars and students mean that dictionaries remain virtually unemployed in the learning-teaching environment, and they may even be thought of as interfering in this process.

The ALRI project in general seeks to correct these misperceptions by presenting the dictionary as a language development tool. Over the years, ALRI researchers have shown that effective dictionary use is fundamental in the domain of social communication. In education, it is essential as a pedagogical tool (Townsend Coles 1988). However, ALRI studies have indicated that there is a general apathy towards dictionaries and a reluctance to their use, that may arise from the teachers' inadequate knowledge of and skills in applying dictionaries in the language class. This apathy and reluctance can also be ascribed to the methods of teaching. Most of the communicative methods used in schools require the deductive approach, where the teacher presents and explains all the material. Also, many language-teaching methods are grammar rather than vocabulary

oriented, so that a dictionary is consulted only when a specific terminological problem arises in class.

5. How is ALRI helping in the development of minority languages?

ALRI now also facilitates the compilation of dictionaries in Zimbabwe's minority languages. Minority mother-tongue speakers are trained to do the required linguistic field work and documentation. Lexicography is used as a step in the development and standardization of these minority languages. Chimhundu (2005) sees the compilation of dictionaries as "a strategy for empowerment of the indigenous language communities in the post-colonial era". He considers lexicography as "the meeting point of all the disciplines that have a bearing on language success in raising the status of a language". Lexicographical research at ALRI is therefore part of the initial efforts to develop and standardize languages which have never been seriously studied before.

5.1 What is being done now?

In 2001, ALRI started collecting data for a corpus from Shangani-speaking areas. The collection of Shangani texts is in progress. Mother-tongue Shangani speakers are hired to do fieldwork. They record oral interviews covering different issues of life. These interviews are transcribed by Shangani mother-tongue speakers who have some basic linguistic background, using diverging orthographies.

There also is a Shangani reference group which comprises Shangani mother-tongue speakers from all areas where the language is used. From time to time, workshops and seminars are held with this reference group to obtain views on methods of developing the language. Plans are in progress to engage the Ministry of Education, Sport and Culture as well as other interested stakeholders to sample the work that is being done so that there is acceptability of the results.

5.2 What are the experiences?

It proves a very difficult task to start with a dictionary of a language in which no previous lexicographical work has been done. Except for the fact that there is no valuable written material in Zimbabwean Shangani, as has been indicated earlier, there is also no orthographic tradition. As a language that has not been well researched, all sorts of grammatical and semantic problems occur. These must all be considered before the actual compilation of the dictionary begins. The greatest challenge faced by ALRI is that of producing a useful dictionary meeting lexicographical standards with the limited available resources as far as personnel, finances and time are concerned.

A major problem is the lack of staff at ALRI to undertake the compilation of the dictionary. There is only one key person possessing the relevant lexicographical training, who encodes the material written in the diverging orthographies the Shangani mother-tongue speakers use when transcribing the interviews.

Except for the inadequate funding to effectively carry out research, ALRI also does not have suitable transport to use when visiting some of the remote areas for data collecting. Furthermore, there are accessibility problems. Shangani-speaking areas are often undeveloped border areas where there are no road networks. Researchers have to walk long distances when carrying out interviews.

Experience at ALRI has shown that lexicographical work is subject to time pressures. Data collection, data processing, and data capturing are necessary but laborious procedures. For a language such as Shangani that has no valuable written material, the work is greatly hindered by time limits. The fieldworkers undertaking data collection must first employ informants and then check, analyze and systematize the data collected by them so that it can eventually be captured at ALRI.

5.3 The nature of the Shangani community

Data collection for corpus expansion is still in progress. The nature of the Shangani community, however, causes problems. This is mainly experienced in some remote areas where the language is spoken. Although mother-tongue speakers, who are fluent in the language do the field work, the interviewees are at times reluctant or uncomfortable to be frank. As a closed community, the Shangani do not want to disclose some of their ways of life, especially some peculiar cultural practices. Fearing exploitation, they still doubt that the work is a genuine linguistic study without any political connotations.

5.4 The orthography and status of Shangani

Like other Zimbabwean minority languages, Shangani does not have a standard orthography. There are many divergences and inconsistencies in the material collected for corpus building at ALRI. These orthographic divergences and inconsistencies mainly result from the fact that most of the written material in Zimbabwe's minority languages consist of recordings of oral interviews taken down by fieldworkers or informally composed texts written by literate speakers, both groups using divergent orthographies. There are therefore no standard texts that can be accessed for the sake of reference. Since minority languages were not taught up to higher levels in schools, publishers were unwilling to invest in them because it was unprofitable to do so. With all this in mind, the researchers at ALRI have to find ways of dealing with the issue of the

Shangani orthography since it is "possible to speak of lexicography only when a writing convention has been developed in a language" (Mathumba 1999: 254). However, language standardization is not a quickly solved issue. There are many channels that need to be followed. Mosel (2002: 3) argues that although the orthographic standardization of a language can be difficult, often being a political matter, it should not retard the compilation and production of a dictionary by incessant debates. What is important is that those concerned should agree on alternative spellings for some words. The decision on which orthography to use will hopefully be arrived at through workshops with the Shangani reference groups as well as other interested stakeholders. Knowledgeable phoneticians and phonologists will be engaged to assist in finding a viable orthographic convention for Shangani. The intention is that the proposed dictionary will become a major normative reference work, which can contribute to developing a standard orthography for Zimbabwean Shangani as well as stimulating the production of more works in and on the language.

Connected to the issue of orthography, is the debate among members of the reference group about the status of Shangani. They are divided on whether to treat Shangani as an independent language from Tsonga or just as a Tsonga dialect. Those who argue that Shangani is a Tsonga dialect feel that it is needless to waste money and resources on issues of standardizing the language. Advocating the harmonization of Shangani and Tsonga spoken in neighbouring countries, they want to adopt the South African or Mozambican Tsonga orthography. However, they agree with those who prefer a unique Zimbabwean orthography, that Shangani now has some peculiar features from Tsonga although they believe that these features can be harmonized. Those who want a different orthography from the Tsonga of neighbouring countries, argue that Shangani is no longer a Tsonga dialect, but an autonomous language with its own dialects.

6.1 Who will be the users of the dictionary and what kind of dictionary should be planned?

The dictionary is first and foremost meant for the indigenous Zimbabwean speakers of Shangani. However, it is seen as having an important role in bringing the language on the agenda for modern development.

Mosel (2002: 4) states that compilers should have to be clear on what exactly is the purpose of their dictionary. It is not just enough to have a dictionary in order to document the lexicon of the language for future generations. The compilers of the Shangani dictionary will have to answer such questions as: For whom is the dictionary primarily intended? Should it be bilingual or monolingual? Does the status of Zimbabwean Shangani warrant the compilation of a monolingual dictionary? Should Shangani follow the way of Shona and Ndebele lexicography? All these questions are some of the challenges the researchers are facing.

The planning and compiling of modern dictionaries demand that specific objectives should be followed. Most important is to identify the dictionary's prospective users and determine the purpose for which it will be used. The answer to this question will also determine what kind of information should be given in the dictionary. So far there is still debate on the type of dictionary to be compiled. There are some who advocate the compilation of a bidirectional Shangani–Shona dictionary while others are for a bidirectional Shangani–English dictionary.

It is argued that since Shangani speakers are found in Shona-speaking areas, a Shangani–Shona dictionary will also help Shona speakers who want to learn Shangani. Since Shangani speakers are a minority, this will also ensure that the dictionary will reach a fairly large number of users.

Those who favour a Shangani–English dictionary base their argument on the fact that such a dictionary will have a much larger number of users than a Shangani–Shona one. This is mainly because speakers of languages like Ndebele will also be able to use it. Actually, all literate people in Zimbabwe will be able to employ it for learning Shangani. The Shangani–English proposal also has a cross-border perspective, because of Shangani being spoken in South Africa and Mozambique too. As a linguistic document, it will indeed add to the mass of lexicological data on African languages already available.

There are also others who feel that the trend at ALRI is the compilation of monolingual dictionaries, so the proposed dictionary should be a monolingual one. But the question with this view is: Does the status of Zimbabwean Shangani in so far as its development is concerned, allow for the compilation of a monolingual dictionary? Should Shangani follow the way of Shona or Ndebele lexicography? Shona and Ndebele lexicography did not just start by compiling monolingual dictionaries. The first dictionaries for both these languages were bilingual ones, English being one of the languages together with Shona and Ndebele. Because of these bilingual dictionaries, they at least had a starting point. Except for this printed material, there were also electronic databases for these languages available. Both Shona and Ndebele were furthermore standardized.

It becomes evident that the challenges of Shangani lexicological development are on all levels, planning, development and publication. For now the focus is on data compilation. It is hoped that with the modern dictionary software that ALRI uses, the final publication will be determined by demand, and will be specifically directed at assisting in the development of other language processes.

6.2 What kind of dictionary would be possible?

As has been indicated above, the Shangani project is constrained by limited resources of money, staff and time but it must be organized in such a way that a useful dictionary is produced. With a view to this critical situation, ALRI is

planning a small comprehensive general Shangani dictionary which aims to meet lexicographical standards. Word lists will be translated from English and Shona. Some will be extracted from a text corpus. The corpus material collected in the documentation project will cover a wide range of different speech situations. However, even with these transferred resources from English and Shona, the work is demanding.

The development of a Shangani dictionary is unique. While historically the priorities of the colonialists for language development did not favour minority ethnic languages, researchers at ALRI are now technically in a better position to creatively and proactively address these issues through a well thought through and scientifically proven process within the promotion policy for the national Zimbabwean languages. In this regard, the following aspects need urgent attention:

- (a) Dictionaries compiled of languages selected for the cultural systems development are essential as they can be instrumental in reviving the linguistic and cultural resources of any ethnic community.
- (b) The development of ethnic languages with the aid of a dictionary will ensure the positive democratization and promotion of social equity and harmony in education, and indeed mutual respect among all ethnic groups, facilitating cultural appreciation and sharing by speech communities.
- (c) The dictionary as a storehouse of cultural, or ethno-cultural knowledge should underscore the fundamental need to facilitate the speech community to fully attain access to all its linguistic and cultural expressions.
- (d) An appropriate compilation of commercializable dictionaries will make the neighbouring communities reach their inner soul, and indeed enhance their cultural values even across borders.

The completion of the first Shangani dictionary will set the basis and the pace for these accomplishments.

7. Conclusion

The facts behind the dictionary development issues within ALRI that may arise when debating the need for an integrative and ethnic-directed cultural policy have been dealt with. These have been raised within the broader issues of the language rights to education of marginalized ethnic communities (cf. Townsend Coles 1988), and, most important, the democratic rights of all citizens to express their culture and also to be taught in their own languages, and indeed at such a level as will be deemed appropriate for the national interests (UN-ESCO 1997). Although the discourse and debate over these issues have often been sensationalized and controversialized by intellectuals and politicians,

mainly in the state bureaucracies, ALRI has chosen a way that will guarantee a producible dictionary even with meagre means.

Undertaking the lexicographical development of Shangani will entail greater justification for its inclusion in the education system of Zimbabwe, at least at elementary levels. Indeed, the lexicographical product will enhance literacy development on at least two levels: orthography and vocabulary. Dictionary resources are therefore very important for minority languages. However, from the discussion, it is evident that there are many factors with which to grapple, chief among them being the financial ones. Human resources can be more easily provided for if the means are available.

References

- Chimhundu, H.** 2005. *Lexicography and Language Raising*. Public lecture presented as part of the Golden Jubilee Celebrations of the University of Zimbabwe, Harare, 13 April 2005.
- Chitiga, M.** 1995. *Codeswitching in the Classroom*. Unpublished M.A. Thesis. Harare: University of Zimbabwe.
- Hachipola, J.S.** 1998. *A Survey of the Minority Languages of Zimbabwe*. Harare: University of Zimbabwe Publications.
- Mathumba, D.I.** 1999. The State of Xitsonga Lexicography. *Lexikos* 9: 254-257.
- Mosel, Ulrike.** 2002. Dictionary Making in Endangered Speech Communities. Wittenburg, P., W. Peters and S Drude (Eds.). 2002. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC2002), Workshop on Tools and Resources in Field Linguistics, University of Las Palmas de Gran Canaria, 26–27 May 2002*. Paris: European Languages Resources Association.
- Townsend Coles, E.K.** 1988. *Let the People Learn*. Manchester: University of Manchester.
- UNESCO.** 1997. *CONFINTEA. Adult Education. The Hamburg Declaration. The Agenda for the Future*. Hamburg: UNESCO.
- Viriri, A.** 2003. *Language Planning: The Conservation and Management of Indigenous Languages in Zimbabwe*. Unpublished paper presented at the ICOMOS 14th General Assembly and Scientific Symposium held at Victoria Falls, Zimbabwe, 27–31 October 2003.
- World Commission on Culture and Development.** 1995. *Our Creative Diversity*. Paris: UNESCO.
- Zimbabwe Constitution.** 1987. Zimbabwe Education Act of 1987, Chapter 55.
- Zimbabwe School Examination Council (ZIMSEC).** 2006. Circular of February 2006.

Diglossic Principles in Shona Lexicography

Esau Mangoya, *African Languages Research Institute (ALRI),
University of Zimbabwe, Harare, Zimbabwe (emangoya@arts.uz.ac.zw)*

Abstract: This article looks at diglossia in Shona dictionaries. Diglossia is seen here as the differential and preferential use of a language or a dialect in dictionaries. The main language spoken in Zimbabwe is Shona. Over a period of about one and a half century, at least twenty Shona lexicographic publications including glossed word lists and medium-sized dictionaries have appeared, examples of which are used in this article to demonstrate the diglossic principles in Shona lexicography. As is the case with many languages, Shona is also in contact with other languages, among them English, which during the colonial era was declared the official language of the country. As a result the dictionaries that were compiled were bilingual. In these dictionaries, involving English and Shona, the bias was towards satisfying the needs of English-speaking users, as much of the detail, including the headword entries, was in English. However, with the need to empower the indigenous languages of the country, there was a deliberate move to promote Shona. This saw a change from the compilation of bilingual dictionaries to monolingual dictionaries which resulted in a different mode of diglossia as the dialects within Shona itself do not enjoy equal status in the dictionaries, because some dialects are given preference above others. This article traces the diglossic tendencies among languages and within the same language in dictionaries involving Shona.

Keywords: DIGLOSSIA, LEXICOGRAPHY, SHONA, SOCIOLINGUISTICS, HIGH VARIETY, LOW VARIETY, SYNONYM, PHONOLOGY, ENTRY, TONE, DEFINITION

Opsomming: **Diglossiese beginsels in Sjonaleksikografie.** Hierdie artikel beskou diglossie in Sjonawoordeboeke. Diglossie word hier gesien as die diskriminerende en begunstigende gebruik van 'n taal of 'n dialek in woordeboeke. Die hooftaal wat in Zimbabwe gepraat word, is Sjona. Oor 'n tydperk van ongeveer een en 'n half eeu, het ten minste twintig Sjona-leksikografiese publikasies verskyn, insluitende geglosseerde woordelyste en middelgroot woordeboeke, voorbeelde waarvan in hierdie artikel gebruik word om die diglossiese beginsels in Sjonaleksikografie te toon. Soos die geval is met baie tale, is Sjona ook in kontak met ander tale, onder hulle Engels wat gedurende die koloniale tydperk as die amptelike taal van die land verklaar is. Gevolglik was die woordeboeke wat saamgestel is, tweetalig. In hierdie woordeboeke waarby Engels en Sjona betrokke was, was die neiging om die behoeftes van Engelssprekende gebruikers te bevredig omdat baie van die besonderhede, insluitende die trefwoordinskrywings, in Engels was. Met die behoefte egter om die inheemse tale van die land te bemagtig, was daar 'n doelbewuste beweging om Sjona te bevorder. Dit het 'n verandering gebring weg van die samestelling van tweetalige woordeboeke na eentalige woordeboeke, wat gelei het tot 'n ander vorm van diglossie omdat die dialekte binne Sjona self nie gelyke status in die woordeboeke geniet nie, maar aan sommige voorkeur bo ander gegee word. Hierdie artikel spoor die diglossiese neigings na tussen tale en binne dieselfde taal in woordeboeke waarby Sjona betrokke is.

Sleutelwoorde: DIGLOSSIE, LEKSIKOGRAFIE, SJONA, SOSIOLINGUISTIEK, HOË VARIËTEIT, LAE VARIËTEIT, SINONIEM, FONOLOGIE, INSKRYWING, TOON, DEFINISIE

Introduction

This article seeks to explore how the principles of diglossia have had a bearing on the production of lexicographic works in Zimbabwe. The linguistic variation to be considered will be both within the same language and between two distinct languages.

The term *diglossia*, coined by Ferguson (1959), has been used primarily in sociolinguistic studies. It refers to the differential use of two language varieties in a given language community and in a given geographical region. In principle, Ferguson has a situation in mind where two or more varieties of the same language are used by some speakers in a speech community under different conditions. This entails some level of switching between the varieties in relevant situations. The concept assumes that there is use of one language variety in one set of circumstances and another under different conditions. The varieties are classified as either being High (H) or Low (L). The H variety is considered superior and the L variety inferior.

Whilst Ferguson's notion of diglossia concerns varieties of the same language, observations in other language situations have led to modifications to Ferguson's mode of diglossia. Fishman's modification broadened the concept of diglossia to embody the assignment of different and separate languages existing in a speech community to serve different tasks in society (Fasold 1984: 40). Thus Fishman includes what he calls a separate code with reference to separate languages. With Fishman's modification, separate languages are assigned different roles whereby some have H status while others are relegated to L status. Both Ferguson's and Fishman's emphasis is on the distinctive function of the varieties within the same society. What Fishman did was to broaden the concept but firstly fully accommodating Ferguson's mode which only focuses on the use of varieties of the same language to its application to the use of separate languages. There are no contradictions between the two scholars and as a result Fishman also adopted Ferguson's rubrics of diglossia which distinguish the H variety which must:

- (a) have a special function assigned to it,
- (b) be a prestigious variety,
- (c) have a long history of literary heritage,
- (d) be acquired through formal education,
- (e) enjoys a stable relationship with its L varieties, and
- (f) be standardised by way of formal codification.

Dictionaries are understood to represent words and also to be repositories of the words in a language. Lexicographic works reflect the social environment in which they are created, this environment again shaping these works. Lexico-

graphy has to follow particular systematising principles to achieve qualities such as user-friendliness, marketability, relevance for and acceptability by the users. If the dictionary is seen as a repository of the words in a language, then it captures the living language, including the social stratification of the language and its varieties.

The Colonial History: Headword Inclusion

In Zimbabwe, English became imposed as the official language during the colonial era. As a result of this, it automatically assumed the H status. This led to the local languages being subjugated to the L status as English became the lingua franca in all official spheres and also in the education system. English had already been fully documented by way of dictionaries and grammar books. As far as lexicography was concerned, English was well developed with numerous publications ranging from large advanced dictionaries to small pocket dictionaries. The fact that all these were compiled abroad where English originated gave them unquestionable authority, hence there has been no reviews of these dictionaries by Africans. They already met the standards in terms of Ferguson's rubrics, coming as a complete package. So English became the H with its special function juxtaposed against the undeveloped local indigenous Ls.

However, in time, the need arose to develop the indigenous languages for enhanced communication. This was not for the purpose of empowering the languages but for the purpose of understanding the African idiom as reflected by the initial attempts by white missionaries. They were concerned with the spread of the gospel to the locals, embarking on standardising the indigenous languages. But the language situation remained with English set as the official language. Whatever followed revolved around the established English language. This was also reflected in the lexicography where the Shona headword entries would be explained in English. Again the H English would assist in standardising the undeveloped Shona. Many reasons have been put forward why English would always filter into and play a major role in dictionaries meant not only to standardise Shona but also to assist English speakers to learn Shona. This article does not dispute this fact but whatever the reasons, the resultant situation is diglossic in nature. Indigenous Shona had to be juxtaposed against the standard English.

This can be demonstrated from Barnes's 1932 dictionary using as example the word *-nyima* (to be stingy).

-nyima, *vb. tr.*; to treat meanly, be mean or stingy to; to stint; *-zinyima*, to deny oneself; *-zinyima-zokudya*, to fast; *wandinyima zokudya*, you have stinted me of food, you haven't given me enough food (*i.e.*, what is due).

Similarly, in Hannan's *Standard Shona Dictionary* of 1959 (second edition 1974),

the headword entries are Shona, but the explanations are in English. The impression is created that the standard should have been referred to English which dominates in the whole dictionary rather than Shona which only provides the headwords. The English H is being accorded the higher status and Shona is being standardised via English as demonstrated in the example of the entry *ngwachata* (idiophone of a rattling noise).

ngwachata (LLL) KMZ ideo of Falling (object that rattles; e g bundle of dry reeds). 2. Z of Falling through weakness. [...]

The Shona entry *ngwachata* is followed by an indication of the tones of the three syllables of the word, *ngwa-* + *-cha-* + *-ta* which are all low as shown by the three Ls (take note of the differences between the use of L in tone and diglossia in this article). The tone indication is followed by the bold letters **KMZ** indicating that the word is used in the Karanga, Manyika and Zezuru dialects. The approach then was to indicate in which Shona dialects the word is used. The absence of the bold **Ko** and **N** means the word is not common in the Korekore and Ndau dialects. After the indication of the dialect comes the English definition. In the same entry, appears another meaning given as definition 2. This second meaning of the word is found only in Zezuru as indicated by the bold **Z**. The definition is also given in English.

The dominating language is English. The medium of communication is predominantly English. The information given by the dictionary is for people who understand English. This observation is reinforced by the fact that this dictionary has an English–Shona index at the end where the headword entries are English and again only Shona equivalents are given while the rest of the definitions are in English, as shown in the following example from the index.

cough n: chikosoro 7 KKoMZ. **C** that causes breathlessness: rutaitai 11 Z.
Whooping **c**: chipembwe 7 KM; chipwembwe 7 M.

Despite the dictionary being labelled *Standard Shona Dictionary* the role Shona plays in the dictionary is minimal. It is clear that the target user should be someone who already speaks English.

Subsequent bilingual dictionaries tried to minimise the dominance of English, e.g. *A Basic English–Shona Dictionary* (1975) and *Duramazwi* (1981) both compiled by Dale. The first has English headwords while the second has Shona headwords. Just as in the *Standard Shona Dictionary*, English plays a dominant role in *A Basic English–Shona Dictionary*. Headwords and explanations are presented in English with the Shona equivalents given at the end. The following is an example of how the headword *African* is treated in the dictionary.

African *belonging to Africa or the people of Africa* adj., e.g. tsika dzevatema
(*African customs*)

Fishman as quoted by Fasold (1984: 54) says: "Whilst diglossia can remain for

years non the less diglossic relationships do change and show characteristic signs of change." This type of change is reflected in Dale's second publication *Duramazwi* where Shona is now given a prominent role. Thus this dictionary has Shona headwords, but the bulk of the information comprising the definitions is in Shona. English equivalents are given last. It has been argued that *Duramazwi* paved the way for the publication of Shona monolingual dictionaries as Shona's ability to describe itself is demonstrated. This can be seen from the following example from Dale's *Duramazwi*.

garo 5, mag-6 (**LL**) Izwi rinoreva kuti nzvimbo inogara pasi kana chinhu chagadzikwa. n *base of a thing; buttock, bottom (polite term)*

This treatment shows why the Shona L is deemed to be moving up towards the H status. In earlier dictionaries, the English H dominates. However, with English speakers gaining more knowledge of Shona, they started focusing to a greater extent on the empowerment of the Shona L. Whilst a positive development is noted with the production of *Duramazwi*, Shona is not totally granted freedom from the English H's dominance. In a way, English still has the authority as it follows on the definition despite the use of Shona in the definitions. As long as the dictionaries compiled are bilingual, the H status of English continues to dominate over the Shona L status.

This situation has resulted in a move towards Shona monolingual dictionaries in an attempt to accord Shona the H status, which it cannot enjoy in a field where it has to compete with English. This has however led to the development of new forms of diglossia. As has been pointed out above, diglossic models can also be transformed or modified. New trends have developed where the diglossic situation has now moved to the internal functions of Shona. Once the focus is on Shona the diglossic function of the dialects of Shona has to be considered. As was the case with the languages, the dialects are similarly in competition for diglossic status.

Competing Dialects

Away from the level at which English and Shona are competing for authority, another form of diglossia makes its appearance at the level of the dialects. This is more apparent in the Shona monolingual dictionaries where the Zezuru, Karanga, Manyika, Ndaou and Korekore varieties are also found not to enjoy equal status. As reflected in the treatment of definitions, and synonyms and variants, Zezuru has been shown to enjoy the H status in comparison with the other dialects.

Dialect H and L Stratification in the Dictionaries

With regard to the dialects, the compilers of the Shona monolingual dictionary

had to have a standard for treating the definitions. *Duramazwi reChiShona* (1996), the first ever Shona monolingual dictionary, was followed by the publication of *Duramazwi Guru reChiShona* (2001). When the different definitions in the two dictionaries are considered, it is notable that the rest of the dialects take a second place to Zezuru. At the same time, it becomes clear that the Shona dialects share a lot of vocabulary. Going through the dictionaries reveals that there was a deliberate preference for Zezuru words where alternatives from the other dialects could have been used as well. Examples of such words are *-rumba* (run) from Ndaou, and *-motsa* (grind finely) from Karanga. These words were avoided in the definition fields. The example below shows how the verb *-rumba* was treated in *Duramazwi Guru reChiShona*.

-rumba D itik. Kumhanya.

The abbreviation D stands for "low tone". The indication *itik* means the verb is intransitive. The meaning given is simply "To run", but the verb root for "run" used with the infinitive *ku-* from the Zezuru dialect is *-mhanya*. So, for the actual meaning of the Ndaou verb *-rumba*, the user has to go to the Zezuru term *-mhanya*. The Zezuru entry is given as follows:

-mhanya [-manya] K itik. ... Kana *uchimhanya* unenge uchikanda tsoka nekukurumidza, uchitomuka zvinokuita kuti ukasike kusvike kwauri kuenda. FAN *-rumba, -tizira, -vanga, -dorinya, -wara, -gijima*.

The variant of the entry follows in square brackets. "K" stands for high tone. After the tone comes "itik" which indicates that it is an intransitive verb. Then follows the actual definition, "When you run you will move your feet forward and bounce which makes you to come quickly to where you are going." After the definition follows "FAN" which is a shortening for the verb *-fanana* (is synonymous). Then, in italics, is given the range of synonyms which do not carry definitions separately, starting with the Ndaou *-rumba*. The definition of the Ndaou term is arrived at via the Zezuru entry.

The occurrence in the dictionaries of these entries peculiar to other dialects is a kind of concession. Their existence is recognised by having them as headword entries cross-referenced to their Zezuru equivalents. Their inclusion and cross-referencing to Zezuru mostly result from complaints raised by speakers of the other dialects. These followed the outreach programmes meant to ascertain the response of the target users to the pioneer monolingual dictionaries in their draft copy stages of production. There was an outcry from the Ndaou and Korekore dialect speakers because many of their lexical items were omitted from these dictionaries which purported to be "purely" Shona. As a result, the second monolingual publication *Duramazwi Guru reChishona* was sensitive to these complaints so that many headwords from the other dialects were included. This at least was possible because, as an advanced dictionary, it was able to accommodate many headword entries.

The first monolingual dictionary *Duramazwi reChiShona* had many of these dialect varieties omitted. Whilst it can be argued that the dictionary was a medium-sized dictionary, which could not contain so many entries, the guidelines showed diglossic tendencies as can be seen from the included headwords which mainly belong to the Zezuru dialect.

However, when headwords from the L status dialects were finally included in the much larger and more accommodative *Duramazwi Guru reChiShona*, they themselves still did not receive authoritative status. Whenever they were entered, as demonstrated above, they were referred to Zezuru H. As mentioned above, the Ndau *-rumba* is cross-referenced to *-mhanya*. Nowhere in the dictionaries is it given the authority and autonomy to be defined and described by Ndau vocabulary. Zezuru plays the H role. To this, Chimhundu (1979: 75) says: "One may claim that Zezuru is becoming the de facto prestige-laden standard language, not vis-à-vis the other dialects as such but their sub-dialects or local varieties."

This H status of the Zezuru dialect is also reflected in the treatment of the synonyms from the other Shona dialects. Although dialects share much vocabulary, they always have vocabulary peculiar to them which, in addition to unique phonological articulation, also helps to identify them. These peculiar dialect lexical items were entered in the monolingual dictionaries but were cross-referenced to the Zezuru forms, not being given the opportunity to carry the definitions themselves. Some such cross-referenced lexical items are given below.

Karanga

-svoda /ʃoɖa/
-shara /ʃara/
-nyara /ɲara/
-bwereketa /byereketa/
shangadzi /ʃangadzi/
zvirokwazvo /zirokwazo/

Zezuru

-nyara /ɲara/ (be shy)
-sarudza /sarudza/ (choose)
-neta /neta/ (be tired)
-taura /taura/ (speak)
shanje /ʃanje/ (type of grass)
chokwadi /tʃokwadi/ (truth)

Korekore

murendo /murendo/
manhoko /manoko/
mbani /mbani/

muenzi /muenzi/ (visitor)
manheru /maneru/ (evening)
ndiani /ndiani/ (who is it)

Ndau

upfu /upfu/
bonore /bonore/
nyamashi /ɲamafi/
-petuka /petuka/
-dhana /dana/

uswa /usɥwa/ (grass)
chibage /tʃi bage/ (maize)
nhasi /nasi/ (today)
-dzoka /dzoka/ (come back, return)
-nyara /ɲara/ (feel ashamed)

Manyika

gwanzi /gwanzi/	nzira /nzira/ (road, path, way)
ehunde /ehunde/	ehe /ehe/ (yes)
dima /dima/	mbambaira /mbambaira/ (sweet potato)

These lexical items from the other dialects are not defined using the dialects' vocabulary. They are cross-referenced to the Zezuru H lexical items which are given the authority to carry the definitions. However, there are cases where the H's authority is reduced or neutralised. These are cases where some of the lexical items are peculiar to certain dialects not having a complete equivalent meaning in the Zezuru H. Examples of this are some headword entries from Ndaui shown below.

Ndaui

yaya /jaja/ (elder sister)
taita /taita/ (elder sister)

In the other dialects, there are terms referring to "sister" in general without the specification of being younger or older. The lexical item that denotes "sister" in general is an adopted term *sisi* from English "sister". However, it has had its meaning slightly shifted from that of English. As an adoptive it refers to "elder sister", and the reference could be used by both younger brother or younger sister, whereas *yaya* and *taita* specifically denote an elder sister, as used by a younger sister. It is only under these special circumstances that the L lexical items are given the opportunity to carry their own definitions without having to be referred to the H variety. However, although they escape the diglossic stratification in not being cross-referenced to the Zezuru entries for meaning, they are nevertheless defined using Zezuru vocabulary, Zezuru therefore once more dominating as the H.

The dominance of the Zezuru H also manifests itself in the treatment of variants which are quite common, particularly between the Zezuru and Karanga dialects. It has been noted that Karanga prefers fricatives whereas Zezuru uses nasal + fricative or stop + fricative combinations as shown in the examples below.

Examples where Zezuru has nasal + fricative combinations:

Karanga	Zezuru
zhira /ʒira/	nzira /nzira/ (road, path, way)
zhara /ʒara/	nzara /nzara/ (hunger)
zheve /ʒeʋe/	nzeve /nzeʋe/ (ear)
zhou /ʒou/	nzou /nzou/ (elephant)

Examples where Zezuru has stop + fricative combinations:

Karanga	Zezuru
svazva /ʃaza/	tsvanzva /tʃanza/ (type of wild fruit)

-svedza /ʃedza/	tsvedza /tʃedza/ (sip off)
shanga /ʃanga/	tsanga /tʃanga/ (grain of seed)
shezha /ʃeʒa/	tseza /tʃeʒa/ (yam)

Lexical items from Zezuru and their variant counterparts from Karanga are given as headword entries in Shona dictionaries. The lexical items from Zezuru have H status as they have authority over their Karanga counterparts. Zezuru has the sole mandate to carry the definitions and any other information elaborating the meaning of the entries. Definitions and elaborative information have to be carried by the term belonging to the H variety, all the lexical items from the L varieties therefore being cross-referenced to the Zezuru H variety.

In certain cases in situations where there is shared vocabulary amongst all the Shona dialects, some dialects are subjugated to this L status regarding variants because of being exclusive of the general trend among the dialects. In such circumstances, the particular dialect remains the only one with L status as has been noted with Manyika's and Korekore's preference for the /w/ phonetic element as in *-waka* (build) where all the other dialects would use /v/, thus having *-vaka*. In such cases, Manyika and Korekore remain exclusive, while the other dialects all acquire H status. If numbers are taken into account, Korekore and Manyika remain in the minority. Those elements peculiar to them are disregarded without any proper justification. There is a deliberate move towards this non-recognition of the /w/ phoneme in those dialects, since all the words from the dialects that involve /v/ would then be variants of the Manyika and Korekore forms. This results in a diglossic situation where Manyika and Korekore words with /w/ onsets remain unrecognised in dictionaries. The existence of the lexical items with /w/ phoneme onsets is disregarded for inclusion in dictionaries, but is only realised in the speech of speakers of the dialects. As Trudgill (1983: 114) puts it: "The low variety ... is used in conversation with family and friends, radio serials, political and academic discussions, political cartoons, and 'folk' literature."

Doke's 1931 standardisation recommendations led to this development of the L status of some of the dialects in which certain phonemes were not recognised in the Shona spelling system. He deliberately recommended that some Ndaou and Korekore lexical items should be omitted. Examples of these peculiar to Ndaou are words having the combination bilabial /m/ and voiceless stop /p/ such as *mpatso* (home). These are normally omitted from dictionaries together with words like *muntu* /muntu/ (person) which have the unrecommended combination of /n/ and /t/. As pointed out above, these are used mainly in speech and are not recognised in written form, which is always the case with L varieties.

A similar case is the nouns of class 11 in Korekore. Korekore has a different realisation of the class 11 prefix *xu-* as in *xukomana* (small boys) and *xumbwa* (small dogs). In the rest of the Shona dialects, these two are realised as *tukomana* and *tumbwa* respectively. The use of /x/ as a prominent feature only in

the Korekore dialect has not been recognised in Shona dictionaries. Despite this exclusion of the Korekore lexical items, some of the dialects managed to challenge the Zezuru H and imposed their own peculiar character. An example is Karanga where the *svi-* prefix as in *svimbudzi* (thin goat) and *svisikana* (thin small girl) is a peculiarity. However, the Karanga dialect receives this kind of temporary H status from Doke's recommendations of standardisation where a certain preference for the Karanga dialect was given.

Leakage from H into L in Adoptives

There is a certain realisation of the H role English has always played and continues to play worldwide. English is considered a language of wider communication. Most technological terms are introduced through English. In the first place, as has been demonstrated above, the purpose of creating the bilingual dictionaries was to raise Shona from its L status. The general lexicographic idea was to bring it on par with English so that it could be used in all spheres of life where at the moment it has limitations. As a result, there has been a deliberate move to introduce as many as possible adopted new technological terms, a move which can be viewed as assisted leakage. *Duramazwi Guru reChiShona* (2001) and the biomedical dictionary *Duramazwi reUrapi neUtano* (2004) made it a deliberate policy to have the technological terms adopted from the English H included, resulting in the listing of terms such as:

kombiyuta (computer)
temberecha, temburicha (temperature)
abhakasi (abacus)

This was done to raise the functional capability of the Shona L as close as possible to the English H. In certain instances, these leakages are viewed with some resistance as the speakers feel that the indigenous L should have some linguistic regeneration. Through some form of affirmative action, they disregard the adoptive leakages and coin their own words to create a feeling that the L is gaining H status: "These are generally used only by those with some special interest in the language who therefore consciously avoid using foreign words when they can help it" (Chimhundu 1979: 78).

Such situations lead to a grey diglossic stratification where the H status of English is partly ignored and Shona coinages are given prominence. However, experience has shown that users, except in formal written language, avoid these coinages in general conversation. Examples of such coinages are:

Coinage	Literal translation
bhizautare	metal horse (bicycle)
hambautare	metal transporter (motorcar)
dutavanhu	human transporter (bus)
mugaradzakasungwa	ever ready person (policeman)

Some of these actually have a somewhat exaggerated denotation. Despite the knowledge that they are not so commonly used, lexicographers always include them in dictionaries to reflect the reality of the historical development of Shona. They were coined in reaction to the diglossic positioning of Shona to English.

Conclusion

Diglossia has been shown to be some form of natural power stratification among languages or dialects used by the same community. There are many factors causing it, including the historical development of the languages and their dialect varieties. The diglossic positioning of the dialects is not so much of a conflict as there is some form of standardised stratification. This diglossic situation develops over a long period of time so that other language-related or -based forms also follow the already existing stratification, as has been shown to be the case with Shona lexicography.

Bibliography

- Barnes, B.H.** 1932. *A Vocabulary of the Dialects of Mashonaland in the New Orthography*. London: The Sheldon Press.
- Chimhundu, H.** 1979. Some Problems Relating to the Incorporation of Loanwords in the Lexicon. *Zambezia* 7(1): 75-91.
- Chimhundu, H. (Ed.)**. 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Dale, D.** 1975. *A Basic English-Shona Dictionary*. Gwelo: Mambo Press.
- Dale, D.** 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.
- Doke, C.M.** 1931. *Report on the Unification of Shona Dialects Carried Out under the Auspices of the Government of Southern Rhodesia and the Carnegie Corporation*. Hertford: Stephen Austin.
- Fasold, R.** 1984. *The Sociolinguistics of Society*. Oxford: Blackwell.
- Ferguson, C.A.** 1959. Diglossia. Hymes, D. (Ed.). 1964. *Language in Culture and Society*: 429-439. New York: Harper and Row.
- Fishman, J.A.** 1971. Societal Bilingualism: Stable and Transitional. Fishman, J.A. 1971. *Sociolinguistics: A Brief Introduction*: 73-90. Rowley, MA: Newbury House Publishers.
- Hannan, M.** 1959. *Standard Shona Dictionary*. Salisbury: The College Press.
- Hannan, M.** 1974. *Standard Shona Dictionary*. Second edition. Salisbury: Rhodesia Literature Bureau.
- Mpofu, N. et al. (Eds.)**. 2004. *Duramazwi reUrapi neUtano*. Gweru: Mambo Press.
- Trudgill, P.** 1983. *Sociolinguistics: An Introduction to Language and Society*. Revised Edition. London: Penguin.

The Role of the African Languages Research Institute in Addressing Language of Instruction Dilemmas in Zimbabwe

Jesta Masuku, *Department of African Languages and Culture, Midlands State University, Gweru, Zimbabwe (jestamasuku@hotmail.com)*

and

Finex Ndhlovu, *School of Languages, Cultures and Linguistics, Monash University, Clayton Campus, Melbourne, Australia (finex.ndhlovu@arts.monash.edu.au)*

Abstract: The lexicographic work of the African Languages Research Institute (ALRI) has played a significant role in attempting to avoid some of the dilemmas associated with using African languages as media of instruction in the Zimbabwean education system. Monolingual Shona and Ndebele dictionaries, biomedical reference works, dictionaries of musical, literary and linguistic terms as well as children's dictionaries constitute part of ALRI's contribution towards the goal of mainstreaming African languages in the education system. This article is an evaluation of the research activities taking place at ALRI. The aim of the article is to demonstrate that if they receive adequate attention through corpus planning, African languages possess the capacity to play an important role as media of instruction across the entire spectrum of the education curricula in Zimbabwe and elsewhere. The article concludes by observing that, if the efforts of ALRI are to succeed, there is need for the co-operation of all stakeholders in language practice.

Keywords: DICTIONARIES, LEXICOGRAPHY, LEXICOGRAPHER, LEXICOGRAPHIC RESEARCH, INDIGENOUS AFRICAN LANGUAGES, AFRICAN LANGUAGES RESEARCH INSTITUTE (ALRI), EDUCATION, CURRICULUM, MEDIUM OF INSTRUCTION, SHONA, NDEBELE, ZIMBABWE

Opsomming: Die rol van die African Languages Research Institute by die hantering van onderrigtaaldilemmas in Zimbabwe. Die leksikografiese werk van die African Languages Research Institute (ALRI) het 'n betekenisvolle rol gespeel om sommige van die dilemmas te probeer vermy wat gepaard gaan met die gebruik van Afrikatale as onderrigmedia in die Zimbabwiese opvoedingstelsel. Eentalige Sjona- en Ndebelewoordeboeke, biomediese naslaanwerke, woordeboeke van musiek-, letterkunde- en taalkundeterme sowel as woordeboeke vir kinders maak deel uit van ALRI se bydrae tot die doelwit om Afrikatale in die hoofstroom van die opvoedingstelsel te plaas. Hierdie artikel is 'n beoordeling van die navorsingsaktiwiteit wat by ALRI plaasvind. Die doel van die artikel is om te toon dat, indien hulle voldoende aandag deur korpusbeplanning ontvang, Afrikatale die vermoë besit om 'n belangrike rol as onderrigmedia oor die hele spektrum van die opvoedingsleerplanne in Zimbabwe en elders te speel. Die artikel sluit

met die waarneming dat, indien die pogings van ALRI wil slaag, daar behoefte is aan die samewerking van alle belanghebbendes in die taalpraktyk.

Sleutelwoorde: WOORDEBOEKE, LEKSIKOGRAFIE, LEKSIKOGRAAF, LEKSIKOGRAFIESE NAVORSING, INHEEMSE AFRIKATALE, AFRICAN LANGUAGES RESEARCH INSTITUTE (ALRI), OPVOEDING, LEERPLAN, ONDERRIGMEDIUM, SJONA, NDEBELE, ZIMBABWE

1. Introduction

The question of developing indigenous African languages for literacy and education has remained a continuing challenge for most postcolonial African nations. Arguments for and against the use of African languages in the school system have been advanced over the years. That there is a need for the use of African languages in all socio-cultural domains and more so in education, has directed and defined paradigms for language policy formulation for many African states. For instance, early in the 1980s, crucial Pan-African conferences on education advanced useful recommendations regarding the issue of mainstreaming African languages in education systems. The 1984 Pan-African Conference on Education held in Yaoundé, Cameroon, on the theme *What School for Africa in the Year 2000?*, critically reviewed the language issue in African education systems. The following year, 1985, the Organization of African Unity (OAU) in the same spirit adopted the *Language Plan of Action for Africa* (see Hartell 1993: v). In this plan, the desire to uplift the institutional and functional status of African languages was emphasized. African member states were encouraged to define their own language policies.

However, it is disappointing that for the majority of African states today, the goal of making these languages viable tools of literacy and education has largely remained mere intention yet to be put into concrete action (King'ei 2001: 122). In the case of Zimbabwe, the African Languages Research Institute (ALRI) has played a leading role in attempting to meet the ideals of the *Language Plan of Action for Africa*. The ongoing lexicographic work at ALRI is one of the pragmatic ways of preparing African languages for wider socio-economic and cultural functions in Zimbabwean national life.

This article, therefore, looks at how the important area of African language lexicography has been placed at the centre of research activities in Zimbabwe. On the part of ALRI, lexicography is viewed as the missing piece in and possible solution to the puzzle associated with the successful use of African languages as media of instruction in African education systems.

For the examination of the contribution of ALRI in addressing the dilemmas regarding the language of instruction in Zimbabwe, this article is organized as follows. The first section gives an overview of the connection between African lexicography and the education curricula. This is followed by a section specifically focusing on how ALRI's activities have so far tried to address the

dilemma of the language of instruction in Zimbabwe. Some of the subthemes covered under this section include an enumeration of completed and ongoing ALRI projects, the dichotomy of lexicography and language policies as well as the role of lexicography in facilitating language acquisition. The article concludes with a statement pointing to future challenges related to the use of African languages as media of instruction.

2. An overview of the potential role of African language lexicography in education

The interplay existing between lexicography, the school curriculum and its subsequent link to the language of instruction in African education systems cannot be over-emphasized. General lexicography concerns itself with the design, use, and compilation of lexicographic reference works as well as the task of evaluating their function and relevance in user communities. African language lexicography goes even a step further in its activities, aiming at the development and documentation of terminologies in various fields, even those fields previously deemed beyond the reach of indigenous languages. In trying to keep pace with scientific and technological advancements, lexicography in the African languages has indeed gone a long way in showing that languages can be manipulated to adapt to the ever-changing scientific and technological advancements in many modern societies. Furthermore, they have shown that indigenous African languages can be promoted to the same level as any other language considered as developed (ALRI Newsletter August 2005: 1). This indeed is an indicator of the strong connection between lexicography and the use of indigenous languages in education. Realizing the interplay that exists between lexicography and the use of indigenous languages as media of instruction in African education systems, Ngara (1982: 133) observes:

[Indigenous] languages should be developed to meet the needs of a modern state and this demands that the vocabulary of these languages be elaborated and modernized ... giving words a scientific orientation, expanding the vocabulary and compiling dictionaries.

With specific reference to Shona, Ngara emphasizes that the use of indigenous languages as media of wider communication is not merely a process of creating appropriate equivalents. These terms should be uniform and be approved by the majority of the users. To this effect Ngara (1982: 135) says:

It is vital that the terms be standardized and be known by the users of the language ... I will recommend that a dictionary of modern Shona terminology be compiled.

The rationale for this is that the use of lexicographic reference works such as dictionaries is an authentic, fast and reliable way of spreading information

about new scientific and technological terms and providing a quick reference for students, translators, media practitioners and others in varying spheres of life. This means that the compilation of lexicographic reference works is indeed a necessary prerequisite for the successful use of African languages as media of instruction in schools. Some of the aspects identified by Ngara such as expanding vocabularies through term creation, developing languages to meet the needs of modern society and ultimately the compilation of dictionaries, happen to lie in the special preserve of lexicography.

The dichotomy of lexicography and education can also be observed in the recommendations made by the delegates to the 1984 Pan-African Conference on Education. In encouraging the development and use of African languages in education, the delegates noted the important role played by African universities in organizing research on these languages. Above all, they recommended the documentation of African languages to make them more functional in society. Obanya (1984: 13) suggested:

The use of African languages for teaching and learning should be intensified and there should be documentation of African languages to make them functional.

That most lexicographic activities in Africa and indeed elsewhere in the world, consist of research carried out by professionals and academics connected to university research units is not a mere coincidence, but an indicator of the strong relationship existing between the disciplines of lexicography and education. In further emphasizing this connection the mission statement of ALRI (Chimhundu 2003: 4) reads as follows:

To research, document and develop Zimbabwean indigenous languages in order to promote and expand their use in all spheres of life.

What emerges from the foregoing is that the development of reference works in the form of orthographies, scientific and technical terminologies and dictionaries constitutes a solid foundation for the successful use of African languages in education. Most important is that these also form part of the essential activities in the field of lexicography. In the following section, the possibility of addressing the dilemma of the language of instruction by using lexicography is further explored.

3. ALRI's contribution to addressing the dilemma of language of instruction through lexicography

This section reviews the lexicographic research projects undertaken by the African Languages Research Institute, discusses their impact on language terminologies, and provides suggestions on how education systems in Zimbabwe and indeed the rest of Africa could take as example its approaches to addressing the problem of language of instruction.

3.1 Background to the research activities of ALRI

ALRI is a non-faculty and non-teaching research institute at the University of Zimbabwe, whose major mandate is to research Zimbabwean African languages with a view to documenting and developing them to promote and expand their overall use (Chimhundu 2003: 4). While major research activities have been in the two national languages, Shona and Ndebele, research has also been extended to minority languages like Kalanga, Nambya and Shangani.

The following is a list of research projects that have been completed, those that are in progress, as well as those planned for the future.

Seven lexicographic research projects, which resulted in the publication of the following dictionaries, have been completed:

- (a) *Duramazwi reChiShona* (1996) — Shona Monolingual Dictionary.
- (b) *Duramazwi Guru reChiShona* (2001) — Advanced Shona Monolingual Dictionary.
- (c) *Isichamazwi SesiNdebele* (2001) — Ndebele Monolingual Dictionary
- (d) *Duramazwi reUrapi neUtano* (2004) — Shona Biomedical Terms Dictionary.
- (e) *Duramazwi reMimhanzi* (2005) — Dictionary of Shona Musical Terms.
- (f) *Isichamazwi SezoMculo* (2006) — Dictionary of Ndebele Musical Terms.
- (g) *Duramazwi reDudziramutauro neUvaranomwe* (2007) — Dictionary of Shona Linguistic and Literary Terms.

Two projects are currently in progress:

- (a) *Isichamazwi SesiNdebele Esilolongwe Sengezwa* (Forthcoming) — Revised and Enlarged Ndebele Monolingual Dictionary.
- (b) *Duramazwi reVana* — Shona Children's Dictionary.

Several projects are planned for the future:

- (a) Ndebele Advanced Monolingual Dictionary.
- (b) Ndebele Biomedical Terms Dictionary.
- (c) Ndebele Linguistic and Literary Terms Dictionary.
- (d) Ndebele Children's Dictionary.
- (e) Shona and Ndebele Mathematics Dictionaries.

It is anticipated that there will also be a focus on more terminological dictionaries in various specialist fields, including health, agriculture, law, commerce, environmental studies, natural sciences and other academic and vocational subjects.

This focus of ALRI is deliberately directed towards ensuring that the African languages of Zimbabwe are developed into relevant media of instruction for use across the entire spectrum of the country's education system.

3.2 Relevance of the research activities to language policy and curriculum issues

A review of the research activities of ALRI in addressing the language of instruction dilemma shows lexicography going beyond its traditional task of dictionary-making into the realm of language planning and development. The question is: How can these activities benefit the education system? The following is a discussion of ALRI's impact on curriculum and language policy issues.

3.2.1 Linking lexicography to the schools' curriculum

The various lexicographic projects in the Zimbabwean indigenous languages have paid close attention to the demands of the schools and higher education curricula. With regard to this, various dictionaries covering terms used in different disciplines offered in Zimbabwean schools, colleges and universities have been compiled or planned. The Shona Children's Dictionary, *Duramazwi reVana*, for instance, targets primary school children falling between the ages of 9 and 13. The main focus of this dictionary has been to select international terms and provide their indigenous language equivalents. The children are subjected to terms used internationally but have the concepts explained in a meaningful language they best understand. Secondary school and tertiary education students will also benefit from the general and specialized dictionaries which treat literary, linguistic, musical, scientific and mathematical concepts, to mention a few of the subject disciplines covered.

While the lexicographers are in the process of compiling these dictionaries, they also have close consultation with the classroom practitioners and curriculum specialists. This enables the lexicographers to obtain simultaneous feedback from the users while the compilation process is still in progress, thus making the reference work an agreed-upon and user-friendly product. With terms captured from the majority of the subject disciplines in the schools' curriculum, lexicography has provided essential references naming and describing varied phenomena in the local languages.

3.2.2 Vocabulary acquisition

Vocabulary acquisition is important in all language teaching and language enhancement initiatives and even more important in the use of these languages as media of instruction in schools. Teachers have to help their students in accumulating a wide vocabulary in using indigenous languages. The dictionary, therefore, becomes one of the most reliable sources for enhancing vocabulary

growth. Once vocabulary growth in the use of African languages has been increased through dictionaries, the ability of using these languages confidently and understanding certain concepts is easy.

3.2.3 Fighting negative attitudes

Attitude problems in the use of indigenous languages in various spheres, even the sphere of education, have a long history. However, these problems are not insurmountable. ALRI's research activities in lexicography have countered the mutual attitudes between the once believed low status of indigenous African languages and the high profile disciplines of science. This is seen in the collaborative efforts made by African language lexicographers and medical professionals resulting in the production of the first Shona Biomedical Terms Dictionary, *Duramazwi reUrapi neUtano*. This dictionary was compiled specifically to help combat communication problems experienced between medical doctors and their patients, some of whom neither speak nor understand English. The Shona biomedical terms presented in this dictionary have undoubtedly enhanced the doctors' communicative competence in using as well as in explaining medical terms in a local language. This resource, although initially targeted at medical practitioners and medical students, can be useful to anyone wanting to use indigenous languages to describe abstract technical and scientific phenomena. Not only do users comfortably express themselves in an indigenous language, but they also shed off their previously held negative notions about what local languages can do and are capable of doing.

Another example of how an indigenous language can be and is capable of being used in the field of science, is the inventory of Ndebele physics terms compiled by an academician at the National University of Science and Technology in Bulawayo. This academician, through collaborative research has provided ALRI with the terms which he uses in his physics lectures. These terms, being standardized and elaborated through seminars held with other science teachers and lecturers, have been incorporated into the Advanced Ndebele Dictionary (AND) database. Furthermore, the same academician has translated his Ph.D. thesis in physics into Ndebele. Hadebe (2004) notes that if teachers and professionals are to change attitudes, they would normally require evidence based on research to convince society of the importance of mother-tongue instruction. What better evidence would the user community require than being shown the extent to which these languages can be and are capable of being used as media of instruction?

The appearance of the first monolingual dictionaries in Shona and Ndebele was received with great enthusiasm. Speaking of the publication of the Ndebele dictionary, Hadebe (2004: 102) observes:

For many Ndebele-speaking people, the publication of the dictionary marked the revival of their language and culture from what is perceived as cultural oppression. The dictionary was considered with national pride.

He further notes that dictionaries must go beyond being mere symbols of national pride and become actual tools of communication. If this is achieved, then dictionaries can be seen as having created a conducive linguistic climate for decisive changes in education policies to be initiated as policy makers would be taking these decisions from an informed position.

4. Conclusion

The commitment of African governments to develop African languages as shown in their blueprint policies has been noted in the article. Although the complexity of the language situation in many countries must be borne in mind, the use of national languages can help in the universalization and democratization of education in Africa. The efforts made in African language lexicography to develop and document the Zimbabwean languages for use in education have also been assessed. The article has shown that any language can eventually be used as a language of instruction provided it has a recognized orthography and its structure is described. For such a language to be suitable for technical and scientific teaching, there has to be stages of fundamental and applied research, and experimentation in term creation leading to the standardization and elaboration of such terms. Fortunately for the Zimbabwean education system, ALRI has already undertaken such research in the field of lexicography. This greatly contributes to addressing the dilemma of the language of instruction in African education systems. However, more work still lies ahead as Brock-Utne (1993: 30), relating the secret behind the success of the Norwegian experience, points out:

In education at the University of Oslo ... none of the books on the required reading list were written in Norwegian. One was Swedish and the rest were English. Together with our professor, we scrutinized difficult English terminology within the professional field of education, found equivalent Norwegian terms and sometimes coined new words. Twenty years later most of the books on the reading list were Norwegian ... It is important that we use our own languages ... and if we stop doing that, the language will not develop and will be less fit as an academic tool.

What is most important now is a complete revision of education policies to acknowledge and reassess the position of African languages as media of instruction.

References

- ALRI Newsletter*. 2005. Award Winning at the 2005 Zimbabwe International Book Fair. August Edition. Harare: African Languages Research Institute.
- Brock-Utne, B.** 1993. *Education in Africa: Education for Self-reliance or Recolonization?* Blindern: University of Oslo.

- Chimhundu, H.** 2003. *Report on the Retreat to Review the ALLEX Project*. Harare: African Languages Research Institute.
- Hadebe, S.** 2004. Improving Dictionary Skills in Ndebele. *Lexikos* 14: 89-104.
- Hartell, R.L (Ed.)**. 1993. *Alphabets of Africa*. Dakar: UNESCO/Summer Institute of Linguistics.
- King'ei, K.** 2001. Challenges and Opportunities in the Development of African Languages in Kenya. Mdee, J.S. and H.J.M. Mwansoko (Eds.). 2001. *Makala Ya Kongamano La Kimataifa: Kiswahili 2000 Proceedings*: 122-133. Dar es Salaam: Chuo Kikuu.
- Ngara, E.A.** 1982. *Bilingualism, Language Contact and Planning: Proposals for Language Use and Language Teaching in Zimbabwe*. Gweru: Mambo Press.
- Obanya, P.** 1984. The Problem of Relevance in the African Schools Programme. *What School for Africa in the Year 2000? Report of the Proceedings of the Pan-African Conference on Education (Yaoundé, Cameroon, 2-9 April 1984)*: 142-165. Morges: World Confederation of Organizations of the Teaching Profession.

*Isichazamazwi SesiNdebele as Reflector of the Moral and Ideological Values of Society**

Nobuhle Moyo, *African Languages Research Institute (ALRI), University of Zimbabwe, Harare, Zimbabwe (nobumoyo@arts.uz.ac.zw)*

Abstract: *Isichazamazwi SesiNdebele* (henceforth the ISN) reflect the moral and ideological values of society. This article analyses the ISN as reflector of the moral and ideological values of the Ndebele society. The ISN highlights fundamental beliefs and ideas of the Ndebele people. The study stems from the observation that the culture of a people will always reflect in a dictionary as culture and language are inextricably intertwined. The aim of this article is to discuss the ways in which the ISN reflects the moral and ideological values of society. The article examines how the society in which the dictionary was created influenced the choice of lemmas and the phrasing of definitions. The focus is on sexist definitions, terms referring to ethnic groups and the treatment of taboo words and religious terms. These terms denote cultural attitudes harboured by the Ndebele society. The article discusses how the treatment of these terms reflects on the values prevalent in the culture of the Ndebele community.

Keywords: MONOLINGUAL DICTIONARY, DEFINITIONS, LANGUAGE, CULTURE, TABOO WORDS, IDEOLOGICAL VALUES, SEXIST ATTITUDES

Opsomming: *Isichazamazwi SesiNdebele as reflekteerder van die morele en ideologiese waardes van die gemeenskap.* *Isichazamazwi SesiNdebele* (voortaan die ISN) weerspieël die morele en ideologiese waardes van die gemeenskap. Hierdie artikel ontleed die ISN as reflekteerder van die morele en ideologiese waardes van die Ndebelegemeenskap. Die ISN belig fundamentele beskouings en opvattinge van die Ndebelevolk. Die studie spruit voort uit die waarneming dat die kultuur van 'n volk altyd in 'n woordeboek weerspieël sal word omdat kultuur en taal onlosmaaklik vervleg is. Die doel van die artikel is om die maniere te bespreek waarop die ISN die morele en ideologiese waardes van die gemeenskap reflekteer. Die artikel ondersoek hoe die gemeenskap waarin die woordeboek tot stand gekom het die keuse van lemmas en die bewoording van definisies beïnvloed het. Die fokus is op seksistiese definisies, terme wat na etniese groepe verwys en die behandeling van taboewoorde en godsdienstige terme. Hierdie terme dui kulturele ingesteldhede aan wat deur die Ndebelegemeenskap gekoester word. Die artikel bespreek hoe die behandeling van hierdie terme waardes reflekteer wat in die kultuur van die Ndebelesamelewing gangbaar is.

* This article was presented as a paper at the Tenth International Conference of the African Association for Lexicography, organized by the Sesiu sa Sesotho National Lexicography Unit, University of the Free State, Bloemfontein, Republic of South Africa, 13–15 July 2005.

Sleutelwoorde: EENTALIGE WOORDEBOEK, DEFINISIES, TAAL, KULTUUR, TABOEWOORDE, IDEOLOGIESE WAARDES, SEKSISTIESE INGESTELDHEDA

1. Introduction

Lexicography is a discipline which exhibits an extensive interconnection between language and culture. Goodenough (1974: 36) points out that "a society's culture consists of whatever it is one has to know or believe in order to operate in a manner acceptable to its members". *Isichazamazwi SesiNdebele* (henceforth the ISN) reflects those cultural aspects that guide the behaviour of members of the Ndebele society. According to Algeo (1990: 2006-2007), "the dictionary can be used as a means of access to the values of the society: it can give a reasonably accurate idea of what is valued, or on the contrary stigmatized" in a particular society. As such the selection and definition of terms in the ISN reflects cultural aspects that are either prized or denounced among the Ndebele. As an authority on usage, the dictionary highlights the social acceptability of certain terms. It can therefore be argued that culture defines the standards of conduct through the language. According to the Whorfian hypothesis, a people's mother tongue offers them a framework for their perception of the world. Thus, while it is imperative in lexicography to know about words and their meanings, there is also a need to know about attitudes, manners and social norms. This might be termed the cultural competence of an individual in general and the lexicographer in particular. Gregerson (1977: 156) states that "language can be studied not only with reference to its formal properties ..., but also with regard to its relationship to the lives and thoughts and culture of the people who speak it". Pride and Holmes (1986: 9-10) echoes similar sentiments:

Any individual member of a community must acquire far more than the formal or structural features of his language(s). He must undergo a process of socialization; he must acquire a knowledge of the social and cultural values of his society, the constraints which the society imposes on behaviour — including language behaviour.

In the light of this statement, it can be said that language forms an important part of a people's definition of life. Through analysing the choice of lemmas and phrasing of definitions, this article aims at discussing the ways in which the ISN reflects the moral and ideological values of the Ndebele society. Focus will be on the treatment of terms displaying the class structure of the Ndebele which include sexist definitions and reference to ethnic groups. The presentation of taboo and religious words will also be discussed. Conforming to a set of beliefs that form the basis of a social system has, however, implications for lexicography in general. This is usually realised in the treatment of "offensive" words where the lexicographer is caught between being forthright and being euphemistic (Chabata and Mavhu 2005) all in the name of upholding the culture of a society.

2. Choice of entries

The choice of entries in the ISN was guided by the frequency-list distribution in the Ndebele corpus. Words lemmatised are those mostly found in the corpus, therefore showing evidence of wide use in both spoken and written Ndebele. The written form of the language, central to language and culture promotion, becomes a permanent source of reference for future generations. In the compilation of a dictionary, a lexicographer therefore has to take cognizance of various cultural aspects, which include certain norms and beliefs embodied in a language. Since language and culture share a complementary relationship, the language used in the dictionary, and the ISN in particular, depicts the worldview of the Ndebele people. What is of significance to this article is the treatment of various entries in the dictionary which will be discussed in relation to the moral and ideological values prevalent in the society.

3. Class structure in Ndebele society

Fromkin and Rodman (1993: 303) mention that the words used to refer to certain individuals or groups reflect "individual non linguistic attitudes" and may also reflect "the culture and views of the society". In Ndebele, various terms indicate the attitudes the society holds against certain individuals. These usually include terms referring to women in general and certain ethnic groups. Thus language is seen as harbouring certain social values promoted by dictionaries, which are repositories of the whole culture. The way the ISN denotes sexist and ethnic attitudes implied by terms and definitions referring to women and foreigners will be discussed in the following two sections.

3.1 Sexist definitions

Kunene and Mulder (1992: 336) point out that "siSwati cultural views about women are reflected throughout the language of siSwati grammar, lexical items and phrases used in everyday speech". This is also true of the Zimbabwean context, in particular among the Ndebele. The entries and definitions in the ISN play a part in the localising and maintaining of women in disadvantaged positions. The language therefore connotes sexist attitudes, which are values in the society. Words referring to certain individuals or certain groups are a reflection of the cultural views of the society because the culture of any society is expressed in its language. Fromkin and Rodman (1993: 312) highlight this:

Just as the use of some words may reflect society's views towards sex or natural body functions or religious beliefs, some words may also reflect racist, chauvinist and sexist attitudes in society.

In Ndebele culture, women, in comparison with men, are usually presented in an unfavourable light. This attitude is reflected in the language. For example:

imitha bz 5. *Imitha* ngowesifazane osewake wazala engendanga. (*This is a woman who has given birth outside of marriage.*)

The definition reflects how Ndebele culture views unmarried women who become pregnant. It carries a derogatory tone since the woman is already viewed negatively on account of her physical condition while the man on the other hand is free to behave at pleasure. In addition to this, the noun *imitha* is derived from the verb *-mitha* (conceive) used in reference to animals, which furthermore conveys a negative view of such a woman. The reason for this treatment is to uphold the traditional view of a woman's morality. Again, a woman whose marriage has failed is referred to as *isaliwakazi*, because society regards her as the offender. Conditions that led to the break-up of the marriage are not considered both ways. The man is not called by an equivalent term to describe his morals. The word is entered as follows in the dictionary:

isaliwakazi bz 7. *Isaliwakazi* ngumuntu wesifazane owehlukana lomkakh. (*A divorcee is a woman who separated from her husband.*)

It carries the notion that it is the woman who divorced the man, not the other way round. Women are thus seen as inferior members of society who have to live within certain parameters set by society. These views are reflected by the dictionary. While women's negative attributes are overemphasized, men's negative attitudes are championed. Thus in Ndebele society there is an idiom which says *indoda yinja* (a man is a dog), meaning that it is not surprising for a man to have a number of women as wives. It is the society which instills these values, every individual being socialized into believing that men are unique members of society who are at liberty to behave at pleasure, while women are set certain measures which restrict their existence in society.

Another of these words which reflects society's view of women is *isifebe* (prostitute) defined as follows in the ISN:

isifebe bz 7. [...] 2 *Isifebe* ngumuntu wesifazana othethweyo oba lobudlelwano bensitha labesilisa angendelanga kubo. [...] (*A prostitute is a married woman who involves herself in private relationships with males to whom she is not married.*)

The word *feba* (commit adultery) is defined in a similar way:

-feba sz gmwa. *Ukufebe* yikuwula komuntu wesifazane owendileyo. (*This is when a married woman commits adultery.*)

The definitions of these words imply that it is only females who behave in a manner unacceptable to society. Men, however, go without chastisement. This cultural attitude towards women indicates male sexist prejudices about women reflected in the language. This sexist language shows that Ndebele culture expects women to behave according to a different and more restrictive code than

men. These cultural views of women evident in the language are also reflected in a lexicographical product.

There are many other names used to describe the conditions in which women find themselves while men on the other hand are not depicted by similar appellations. This points to an ideology in society which is constantly watching the behaviour of women. Thus they are given names in accordance with their physical conditions.

Language tends to reflect sexism in paired words which usually carry some often-noticed overtones. According to Fasold (1993: 113), these words are supposedly semantically equivalent except that the one refers to male and the other to female. The Ndebele language demonstrates that these words have different associations which are reflective of the sexist attitudes of society. For example, in Ndebele an unmarried woman is referred to as *umazakhela* (literally: one who builds for herself) while a bachelor is referred to as *insumbelume*. They are entered as follows in the ISN:

umazakhela bz 1a. Umfazi othiwa ngumazakhela ngosekhulile ongendanga, kumbe owatshiya umendo wazakhela owakhe umuzi. (*A spinster is an old unmarried woman or a divorcee living by herself.*)

insumbelume bz 9. *Insumbelume* ngumuntu wesilisa osedlule ibanga loku-thatha kodwa elokhu engelamfazi. (*A bachelor is a man who is past the rightful age of marrying but is still unmarried.*)

In these cases, Fasold (1993: 113) observes that a bachelor is seen as unmarried by choice and living happily, while a spinster conjures the image of an old, unappealing woman living an unfulfilled life in consequence of her failure to marry, hence the need to fend for herself as reflected by the literal meaning of the Ndebele word *umazakhela*. This also shows that Ndebele society believes that "motherhood is the highest goal of traditional African woman" (Lippert 1972: 166), which can only be achieved in marriage. The language used in the ISN reflects these sexist views and values of a male-dominated society.

The sexist values of Ndebele society are also shown in entries where women are given names on account of their physical conditions. For example, the term *inyumba* is defined as follows:

inyumba bz 9. *Inyumba* ngumuntu wesifazane ongazaliyo. (*A barren woman is one who has failed to conceive.*)

Biology has proven that it is not only women who are barren, as African societies like the Ndebele tend to believe. However, since it is the staunch belief of Ndebele society that only women can be barren, women, not men, bear the negative name. This prejudice of a patriarchal society where women receive the blame results in the stereotyping of women. The ISN thus expresses the ideology of society.

The presentation of Ndebele women in the ISN also indicates the status

society accords them. In general, women have always been relegated to subordinate positions in African societies and this also applies in Ndebele society. It is these underlying societal values that result in the negative depiction of women. The dominance of patriarchal ideology grants a privileged status to males in comparison with females. In Ndebele society, it is therefore not expected to find a man who is unsuccessful in life. Thus such a man is termed *indunayabafazi* (chief of the women), which appears in the ISN as follows:

indunayabafazi bz 9. 1 *Indunayabafazi* ngumuntu wesilisa ongasumuntu walutho. [...] (1 A chief of women is a very useless man. [...])

This indicates the inferior status accorded women in Ndebele society. An unsuccessful man is not only termed a chief (someone with power and authority) but also named in accordance with his superiority over women. This stems from the general patriarchal belief that women cannot be successful in life.

3.2 Ethnic groups

Fasold (1987: 148) observes that "attitudes towards language are often the reflection of attitudes towards members of various ethnic groups". This is also evident in the Ndebele language. During the pre-colonial period, the Ndebele community was socially divisive. Those associated with the royal family of the (abeZansi) Ndebele state enjoyed a high and prestigious social status. The Ndebele language used in the ISN thus mirrors the divisions among the Ndebele community. People not originally Nguni are referred to as *amahole* implying their inferior status as compared to the original Ndebeles from Zululand. This superiority complex in Ndebele culture is reflected in the language and therefore also in the dictionaries. The word *ihole* is defined as follows:

ihole bz 5. *Ihole* ligama lokuthuka ingqe ngumuntu ongasimzansi ngokudabuka. Akulunganga ukulisebenzisa kwabanye abantu. (This is an insulting term used to refer to anyone who is not Nguni by origin. It is not proper to use this term to refer to other people.)

This word is used to mock anyone who is not of Nguni ethnic origin. The definition shows that *ihole* is insulting under all circumstances. This kind of attitude is a result and a manifestation of the underlying values and beliefs which constitute a community's culture. Pride and Holmes (1986: 9) observe that an individual member of a community must acquire knowledge of the cultural values of his society and the constraints which the society imposes on behaviour, including language behaviour.

Similar to *ihole* are the words *inyasarandi* (one who originates from Malawi) and *iTshabi* (one who is not Ndebele by origin) which have the following definitions:

inyasarandi bz 5. *Inyasarandi* ligama elithukayo lomuntu odabuka eMalawi. (This is an insulting term referring to someone who originates from Malawi.)

iTshabi bz 5. Leli ligama lokweyisa elisetshenziswa ngamaNdebele limele wonke umuntu onsundu ongasiNdebele ngomdabuko. (This is a pejorative term used by Ndebeles to refer to anyone who is not Ndebele by origin.)

The above words and definitions reflect the values prevalent among the Ndebele people. Thus different ethnic groups were given names in accordance with their attitudes towards the group. Therefore language attitudes generally reflect ideological or cultural differences among the different ethnic groups. A lexicographer becomes a reflector of the culture expressed by a particular language. The language used in the dictionary therefore mirrors the ideology of the society by which it is spoken.

4. Taboo words

Fromkin and Rodman (1993: 303) are of the view that dictionaries often give clues to social attitudes. In Ndebele culture, certain qualities are vested in the language making it possible to comment on certain situations when using proper discourse. For example, it is taboo among the Ndebele directly to refer to *ukufa* (death of a king). It can only be expressed in euphemistic terms. Nyathi (2001: 124) says: "When a king dies, it is said 'the mountain has fallen' *intaba idilikile*. Sometimes ... 'he has bowed down' *inkosi isikhotheme*." For in Ndebele culture, it is considered disrespectful to say *umuntu ufile* (a person has died). Instead, it is said a person has passed away (*usedlule*) or has sunk (*usetshonile*). This shows that death is a taboo subject among the Ndebele, therefore making it unacceptable to refer to it directly. The lexicographer has to be sensitive to this cultural attitude.

As mentioned earlier, a people's mother tongue offers them a framework for their perception of the world. In many cultures and that of the Ndebele in particular, it is also taboo directly to refer to terms with sexual connotations, as becomes evident from dictionaries. In these cultures, reference to such terms must be done euphemistically. Similarly their treatment in dictionaries is in accordance with what the society expects. Although taboo words were entered in the ISN because of their presence in the Ndebele corpus, the phrasing of the definitions clearly indicates that it is against the social norms of the Ndebele to define them explicitly. Doing so will be a violation of the moral code of conduct on the part of the lexicographers. As a way of upholding the cultural values of the Ndebele people, the editors of the ISN defined these terms in a polite manner. For example the definition of the word *igolo* (vagina) reads as follows:

igolo bz 5 *Igolo* yisitho sensitha sowesifazana esixhumanisa isibeletho lezitho zensitha ezingaphandle. Ligama elihloniphisayo leli. (The vagina

is a female private body part attached to the womb and other private parts on the outside. The term is impolite.)

In this definition, euphemistic expressions like *yisitho sensitha* (private part) are used to explain this word. The definition points out that *ligama elihloniphisayo* meaning 'the term' should be used with prudence. An explicit definition which could have defined the word in terms of its function in childbearing or sexual intercourse is unacceptable in Ndebele culture. Hence the definition reflects the Ndebele moral values. Another definition treated in a similar manner is the word *ubolo* (penis) defined as follows:

ubolo bz 11. *Ubolo yisitho sendoda kumbe inyamazana enduna esise-tshenziswa ukuchema kumbe ukwandisa uhlobo. (The penis is a part of the body of a male being or animal that is used for urinating as well as increasing the species.)*

In this example, the word *ubolo* (penis) is defined only in terms of its use in conception and urination. The euphemistic expression *ukwandisa uhlobo* is meant to fit the definition into the acceptable cultural norms of the Ndebele. This shows that lexicographers have a specific mandate to use language acceptable to the society, because language is inextricably bound up with the culture of a people. Chabata and Mavhu (2005: 255) concur that "the lexicographer's pen is sometimes not free to write its owner's views and convictions". Thus the lexicographer is forced to write in a manner reflecting the sensitivities of the society for whom the dictionary is intended. Another example which supports this point is the way the back matter of the ISN is presented. The back matter covers a variety of cultural information which includes, among others, the names of different body parts. While most of the parts have been entered, those referring to the private parts have deliberately been omitted. However, as shown above, these words have been defined in the central list of the dictionary in a manner acceptable to the society when dealing with taboo words. Lexicographers as members of the society simply conform to a set of beliefs that form the basis of their social system.

5. Religious words

Religious words too are associated with the belief system of the Ndebele. Nyathi (2001: 4) observes that "it becomes very difficult, in fact, impossible, to isolate religion from other aspects of Ndebele culture". The ISN indicates that the Ndebele society basically believes in two types of religion, namely the traditional African religion and Christianity. Various terms relating to both the traditional African and Christian beliefs are entered in the dictionary.

The culture of a society contains all the beliefs that a society has about itself. In traditional Ndebele culture, *amadlozi* (ancestors) are viewed as a source

of livelihood for the family. The basis of Ndebele religion lies in the belief in ancestral spirits. Generally the life of the Ndebele centres on the recognition and appeasement of the ancestors to solicit more protection and blessings from them. The definition of *amadlozi* reads as follows in the ISN:

amadlozi bz 6. *Amadlozi yimimoya yabantu asebafayo okukholelwa ukuthi iyaphenduka izelondoloza kumbe ukuphazamisa impilo zabaphilayo. (Ancestors are the spirits of the dead believed to come back either to take care of the living or to disturb their lives.)*

The use of the word *okukholelwa* indicates that it is not easy to arrive at the underlying values or beliefs in ancestors, hence it recognizes the fact that not everyone in Ndebele society believes in or has faith in ancestral spirits. For someone who strongly believes in the propitiation of the ancestors, these values are highly recognized. The fact that the definition is not authoritative might not satisfy those who believe in the ancestral religious practice. However, the tone of the definition seeks to strike a balance with those who believe in other religious practices, further highlighting the differing values within Ndebele society.

In the same way, the word *uNkulunkulu* is defined as follows in the ISN:

uNkulunkulu bz 1a. *UNkulunkulu kukholelwa ukuthi ngumdali wezinto zonke emhlabeni [...] (God is believed to be the creator of everything on earth.)*

This example also shows that it is difficult to make the definition authoritative, because it refers to something that cannot be proven. However, Christians might feel that the definition does not accord God the rightful status as it carries the notion of impossibility of proof. The *Random House College Dictionary* (1988) enters *God* in this way:

God [...] the Supreme Being, the creator and ruler of the universe.

This definition, in contradistinction to the one in the ISN, expresses the indisputability of God's part in the existence of the universe. Perhaps this difference stems from the history of Christianity among the Africans in general and the Ndebele in particular. Before the advent of Christianity, the Ndebele used to communicate with God via the ancestors and some still do even today. However, the introduction of Christianity among the Africans brought with it some negative views on beliefs in the traditional ancestral religion. In a 2003 interview with Hadebe, the editor-in-chief of the ISN, he pointed out that in defining most of these terms associated with the belief systems of the Ndebele, the editors felt that Pelling in his bilingual English–Ndebele dictionary had suppressed the African religion, thus they had to shed more light on it when defining words associated with it. For example, Pelling (1971) enters the word *isangoma* in the following way:

is-angoma [...] witch doctor.

This definition presents *isangoma* (spirit medium) in an unfavourable, derogatory light where an association with witchcraft is implied. It reflects the belief of Christians that the spirit medium is inspired by the devil. However, because some Ndebeles believe in spirit mediums, it is entered in a different manner, with no undertones of witchcraft, in the ISN. In sharp contrast to Pelling's definition, it is presented as follows:

isangoma bz 7. *Isangoma* ngumuntu oledlozi lobungoma njalo okuvumiso. (A spirit medium is a person who possesses the spirit of the ancestors which helps in diagnosis.)

Unlike Pelling, who describes *isangoma* as a witch doctor, this definition connects it with the belief in ancestral spirits. Not everyone in Ndebele culture is a Christian, therefore words pertaining to both belief systems of the Ndebele should be entered and defined explicitly in the dictionary, the reason being that a dictionary's greatest value is to give access to the full range of a language. The ISN therefore mirrors the various religious beliefs in Ndebele society.

6. Conclusion

This article has discussed the different ways in which language reflects the moral and ideological values of a society. As the reflectors of the culture of a people, lexicographers are compelled to be truly descriptive and exactly explain what the words mean in a given society. While this is achievable with other word categories, taboo subjects have always constrained lexicographers into conforming to a set of beliefs forming the basis of a social system. As a result taboo words are dealt with in a manner that does not arouse controversy in a specific society. This obviously has implications for lexicography as a discipline that considers meaning as central. However, this clearly shows how the culture of a society permeates into all spheres of life including dictionary compilation.

References

- Algeo, J. 1990. American Lexicography. 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*: 1987-2009. Berlin: Walter de Gruyter.
- Chabata, E. and W. Mavhu. 2005. To Call or Not to Call a Spade a Spade: The Dilemma of Treating 'Offensive' Terms in *Duramazwi Guru reChiShona*. *Lexikos* 15: 253-264.
- Fasold, R. 1987. *The Sociolinguistics of Society*. Oxford: Blackwell.
- Fasold, R. 1993. *The Sociolinguistics of Language*. Oxford: Blackwell.

- Fromkin, V. and R. Rodman.** 1993³. *An Introduction to Language*. New York: Harcourt Brace Jovanovich College Publishers.
- Gregerson, E.A.** 1977. *Language in Africa: An Introductory Survey*. New York: Gordon and Beach.
- Goodenough, W.H.** 1974. Cultural Anthropology and Linguistics. Hymes, D. 1974. *Language in Culture and Society*: 36-39. New York/Evanston/London: Harper and Row.
- Hadebe, S. (Ed.).** 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Kunene, E.C.L. and J.G. Mulder.** 1992. Linguistic Considerations of Some Cultural Attitudes in siSwati. Herbert, R.K. (Ed). 1992. *Language and Society in Africa: The Theory and Practice of Sociolinguistics*: 335-344. Johannesburg: Witwatersrand University Press.
- Lippert, A.** 1972. *The Changing Images of Women as Viewed in Literature of English and French Speaking West Africa*. Unpublished Ph.D. Thesis. Bloomington: Indiana University.
- Nyathi, P.** 2001. *Traditional Ceremonies of AmaNdebele*. Gweru: Mambo Press.
- Pelling, J.N.** 1971. *A Practical Ndebele Dictionary*. Revised Edition. Salisbury: Longman Rhodesia.
- Pride, J.B. and J. Holmes (Eds.).** 1986. *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin Books.
- Stein, J. (Ed.).** 1988. *Random House College Dictionary*. Revised Edition. New York: Random House.

The User Perspective in Lexicography: The Lemmatisation of Fixed Expressions in *Duramazwi Guru reChiShona*

Nomalanga Mpofu, *African Languages Research Institute (ALRI),
University of Zimbabwe, Harare, Zimbabwe (nomalanm@yahoo.com)*

Abstract: The article discusses the user perspective and information retrieval in relation to the lemmatisation of specific multi-word lexical units, namely fixed expressions, in the Shona monolingual dictionary, *Duramazwi Guru reChiShona*. It shows that the decisions arrived at in lemmatising fixed expressions were influenced by a user-driven approach. The article gives a comparative analysis of how fixed expressions were treated in previous Shona dictionaries and how they were subsequently dealt with in *Duramazwi Guru reChiShona*. Previous dictionaries have grappled with the problem of giving fixed expressions as run-on entries. Against the background of the user perspective, it will be argued that the lemmatisation of fixed expressions in monolingual dictionaries has certain advantages over previously used strategies.

Keywords: FIXED EXPRESSION, HEADWORD, IDIOM, INTENDED USER, LEMMA, LEMMATISATION, MACROSTRUCTURE, MICROSTRUCTURE, MULTI-WORD LEXICAL UNIT, PITHY SAYING, PROVERB, RUN-ON ENTRY, USER, USER-FRIENDLY, USER PERSPECTIVE

Opsomming: Die gebruikersperspektief in die leksikografie: Die lemmatisering van vaste uitdrukkings in *Duramazwi Guru reChiShona*. Die artikel bespreek die gebruikersperspektief en inligtingsherwinning met betrekking tot die lemmatisering van spesifieke meerwoordige leksikale eenhede, naamlik vaste uitdrukkings, in die Sjona-eentalige woordeboek *Duramazwi Guru reChiShona*. Dit toon dat die besluite waartoe gekom is by die lemmatisering van vaste uitdrukkings beïnvloed is deur 'n gebruikersgedrewe benadering. Die artikel gee 'n vergelykende ontleding van hoe vaste uitdrukkings in vorige Sjonawoordeboeke bewerk is en hoe hulle vervolgens in *Duramazwi Guru reChiShona* behandel is. Vorige woordeboeke het met die probleem geworstel om vaste uitdrukkings as deurloopinskrywings te gee. Teen die agtergrond van die gebruikersperspektief word aangevoer dat die lemmatisering van vaste uitdrukkings in eentalige woordeboeke sekere voordele het bo vroeër gebruikte strategieë.

Sleutelwoorde: VASTE UITDRUKKING, BEDOELDE GEBRUIKER, DEURLOOPINSKRYWING, GEBRUIKER, GEBRUIKERSPERSPEKTIEF, GEBRUIKERSVRIENDELIK, IDIOOM, LEMMA, LEMMATISERING, MAKROSTRUKTUUR, MEERWOORDIGE LEKSIKALE EENHEID, MIKROSTRUKTUUR, PITTIGE GESEGDE, SPREEKWOORD, TREFWOORD

1. Introduction

The article discusses aspects of information retrieval in a monolingual dictionary against the background of the user perspective. In order to show that lemmatising specific multi-word lexical units (MWLUs), namely fixed expressions has the advantage of creating a more user-friendly dictionary as opposed to entering them as run-on entries, it analyses the macrostructural treatment of fixed expressions in *Duramazwi Guru reChiShona* (2001), the advanced Shona monolingual dictionary. It is easier for a user to access information from the macrostructure than to search for it within another entry in the microstructure.

The macrostructure of a dictionary consists of the lemmata arranged according to a specific structure. Burkhanov (1998: 146-147) distinguishes three types of macrostructure in dictionaries. The first type is the ideographic macrostructure where lemmata are organised by semantic affinity, while the second type is the alphabetical macrostructure where lemmata are organised according to the alphabetical ordering of the graphical word. The analogical macrostructure constitutes the third type, involving a combination of both the ideographic and alphabetical ordering. The macrostructure that this article refers to is the alphabetical type. The microstructure of a dictionary, on the other hand, is the internal structure of the dictionary entry itself. For instance, pronunciation, the lexical category of a word, the ordering of senses, examples and synonyms are all part of the microstructure.

By fixed expressions in this article are meant proverbs, idioms and pithy sayings. A proverb is a statement full of wisdom handed down from generation to generation. Didactic in content, it is usually used to advise, admonish or instruct. *Collins COBUILD English Language Dictionary* (1991) defines a proverb as a short sentence that is often quoted, giving advice or telling something about human life and problems in general. Shona proverbs are complete statements having a bipartite structure consisting of two propositions: the first proposition foregrounds the proverb, while the second proposition complements the first. The overall meaning of the proverb is derived from the balance and cross-correspondence between the two propositions. Detailed descriptions of the form of Shona proverbs are given by Fortune (1975–1976) and Chimhundu (1980).

An idiom is a phrase whose meaning cannot be predicted or deduced from the individual meanings of its constituent parts. Its meaning is figurative (Kipfer 1984: 103). As its meaning is greater than the meaning of its constituent parts, it must be considered a unity. Structurally, Shona idioms are verbal in form, always beginning with the infinitive verb *ku-* 'to'.

Pithy sayings in Shona are neither proverbial nor idiomatic in meaning, but are statements used in everyday language to comment on events, or in some cases, to offer advice.

Duramazwi Guru reChiShona (hereafter referred to as DGS) is divided into two sections. Section 1 contains the main A–Z section of the dictionary includ-

ing idioms, while section 2 contains proverbs and pithy sayings. The article will look specifically at the way proverbs, idioms and pithy sayings were lemmatised in DGS with particular focus on fixed expressions as lemmata vs. fixed expressions as run-on entries. The article will show that a user-friendly dictionary takes the needs of the user regarding access to information and retrieval of information in a dictionary into cognisance. It will also indicate that the decision to lemmatise fixed expressions in the DGS was influenced by the user-driven approach.

2. The user perspective in lexicography

The user perspective or the user-driven approach in lexicography refers to the dictionary compiler's recognition of the importance of the needs of the user of the dictionary with reference to the type and function of the dictionary, and the structure of its entries in the macrostructure as well as in the microstructure. The macrostructure is most important as it forms the access point to the information contained in the dictionary. Before a dictionary is compiled, the lexicographer has to identify the potential user of the dictionary and plan the nature and structure of the dictionary according to this target user. It is therefore the target user who determines the type of dictionary to be compiled. The target user also determines the overall macrostructural and microstructural characteristics of the dictionary.

The other important aspect of any dictionary is whether the user is able to find the looked-for information. The macrostructure of the dictionary has to be designed in such a way that information retrieval is effected easily. In other words, the dictionary has to be user-friendly, which means that it has to satisfy the following parameters:

- Macrostructural accessibility: Is the user able to find the looked-for information?
- Information retrieval: Is the information found with the least cognitive effort?
- User-friendliness: Is the information given where the user expects to find it?

The guiding principle of the user-driven approach is that lexicographers should take the needs of the users into consideration during the planning and compilation stages of the dictionary project. A dictionary therefore has to be organised in such a way that it caters for the users' needs. Lexicographers should also not take for granted that all users have the necessary dictionary searching skills. Hence, a dictionary should be compiled in a way that meets the needs of both the skilled and unskilled users, thereby simplifying the information retrieval, which is of paramount importance in any dictionary. Dic-

tionary users have to be able to find what they are looking for in as little time as possible.

Gouws and Prinsloo (1997: 46) state:

User-friendliness in dictionaries implies that the contents of the dictionary are made as accessible to the user as possible. ... The macrostructure remains the main access structure of any dictionary with a strict alphabetical ordering system. Lexicographers too often neglect the importance of a well-designed macrostructure as a functional component of the total linguistic contents of a dictionary ...

Proverbs, idioms and pithy sayings are not only in daily use in speech and conversation but also form part of the Zimbabwean school curriculum. Students are examined not only on grammar, composition writing and comprehension skills, but are also tested for their ability to give the full forms of these fixed expressions as well as their meanings. Students may also be required to write an essay based on a proverb, an idiom or a pithy saying. In order to do this they have to know what a particular proverb, idiom or pithy saying means and how it is used in context. The justifiable assumption is that it is these students who consult a dictionary like the DGS. Meaning and context are the domain of dictionaries, hence the inclusion of fixed expressions in the DGS. By having fixed expressions lemmatised, the dictionary becomes a complete reference manual, where the user can find all the looked-for information in one place, in this way also being time-saving and cost-effective.

3. The treatment of fixed expressions in previous Shona dictionaries

There is no set method for the presentation of MWLUs like fixed expressions in dictionaries. Usually lexicographers use the method that best suits the type of dictionary to be compiled, and the target users in question. The problem faced by lexicographers past and present has been whether MWLUs should be headwords, therefore being part of the macrostructure, or whether they should be treated under one of the single-word lemmata, thus forming part of the microstructure (Svensén 1993: 207). Svensén says that it has been difficult for lexicographers to break away from the principle that a headword is the same as a graphical word. Hence dictionary macrostructures have taken the graphical word as their basis. Gouws (2003: 39) points out that dictionaries have been characterized and dominated by what he refers to as a word-bias. This bias has been so entrenched in lexicography that the norm in dictionaries has always been to enter words and to nest any units larger than words into an entry as run-on entries.

There have, however, been inconsistencies in the way MWLUs have been treated in dictionaries. Čermák (2000: 490) identifies two methods in lexicographers' treatment of MWLUs. One approach is to list them under a specific sense of the single-word lemma. With this method, a proverb like *Chara chimwe hachi-*

tswanyi inda (One finger cannot crush a louse) would be placed under either of the nouns *chara* (finger) or *inda* (louse), or under the verb *-tswanya* (crush). The other approach is to list them separately at the end of the lemma. With this method, one word in the MWLU would also have to be chosen as lemma under which to include it.

This method is also problematic because, in Shona for instance, some phrasal entries can begin with the same word, such as *-bata nekuseri kweruoko* (lit. touch with the back of the hand, i.e. treat badly), and *-bata denga* (lit. touch the sky, i.e. be overcome with joy), *-bata jongwe muromo* (lit. touch the cock on the mouth, i.e. be an early riser). This would mean that an entry such as *-bata* (touch, hold), which in itself has 18 senses in the DGS, would be very congested. In terms of information retrieval, such an entry would not be considered user-friendly as the more congested an entry, the more difficult to retrieve required information in a short time and with the least effort. Possibly such an entry would also be clumsily and awkwardly constructed.

In both methods cited by Čermák, MWLUs are treated as run-on entries under a single-word lemma. Čermák (2000), Moon (1992) and Lorentzen (1996) are critical of both methods as they result in many inconsistencies, leaving the user guessing under which sense an MWLU might have been listed. Consequently, it is difficult to find MWLUs in any reliable way, because, as Čermák (2000: 490) says, this guessing game is rather difficult to follow.

With reference to stem lemmatisation vs. word lemmatisation, Gouws and Prinsloo (2005) point out that in Bantu language dictionaries lemmata are becoming heavier to the left. While this is true with regard to morphological considerations, lemmata are also becoming morphosyntactic. As a result of the changing trends in lexicography, dictionaries are now lemmatising units larger than the word in a bid to meet the needs of different users.

3.1 Proverbs

In previous Shona dictionaries, namely Hannan (1959, 1974) and Dale (1981), proverbs were not given lemma status but were entered as run-on entries under what was considered to be one of the semantically heavy words in the phrase or under the word where the user is most likely to look for it. The general definitions of the word would be given first and then, following that, the other forms or combinations in which the word is found. In Hannan, proverbs were presented as follows:

- (1) **-chenga** [L] KKoZ v t Keep close. 2. Value. > *ruchengera*. *Chenga ose manhanga, hapana risina mhodzi*: keep all the pumpkins, there is none that has no seeds (i.e. be thrifty) prov. [...]
- (2) **zambuko** [LLL] KMZ n 21 Ford, drift, river crossing. < *-yambuka*. *Aiva madziva ava mazambuko, aiva mazambuko ava madziva*: what used to be pools are now crossings, what used to be crossings are now pools (i.e. the old order changeth yielding place to new) prov.

Dale (1981) used an almost similar method to that of Hannan:

- (3) **-CHENGA** t (L) Kunyatsochegeta chinhu. [Chenga ose manhanga; hapana risina mhodzi. Preserve all pumpkins; every one of them has seeds. prov 1447 i.e. People have equal dignity whoever they are.]
- (4) **-ROVA** t (H) Kuranga neshamhu kana chibhakera etc. [...] Charova sei chando; kuzoona hamba yokwira mumuti? How cold must it be that (even) the tortoise is climbing a tree? prov 573 i.e. Hardship can induce extraordinary changes of fortune.]

3.2 Idioms

Idioms were also treated as run-on entries under what was perceived to be the most semantically heavy noun or verb. When they were treated as run-on entries, their placement was problematic, that is, the choice of a suitable noun or verb under which to include them. For instance, the idiom *(ku)tamba nema-dhaka pasina mvura* (lit., to play with mud where there is no water, i.e. to do things that can get you into trouble) can be given as a run-on entry under the verb *(ku)tamba* (to play) or under the nouns *madhaka* (mud) or *mvura* (water). Similarly, *(ku)bata chibharo* (to rape) can either be entered under the verb *-bata* (touch, hold) or under the noun *chibharo* (hard labour). Such an either-or situation can result in a confused user. Zgusta (1971: 276) says:

Apart from the problem of their selection, ... there is the problem of where to list these expressions. Smaller dictionaries list them at the end of the entry ... Bigger dictionaries list idiomatic expressions at the end of the examples of free combinations of words.

Hannan (1959) gave idioms as run-on entries under what he must have considered the most common noun or verb. The idioms are placed last within the entry, after the definitions and usage examples. In some cases, he marked them 'Note idiomatic usages'. For example:

- (5) **-bata** [H] KKoMZ v t Touch. Hold. Catch [...] 2. i KMZ Be strong [...] Bata *imbwa*: catch (or hold) the dog [...] Note idiomatic usages: kubata *maoko*: offer sympathy, commiserate (in bereavement); kubata *mumwe chibharo*: force, rape; kubata *mumwe kumeso*: deceive by flattery, bluff; kubata *pamuviri*: conceive; kubatisa *mumwe pasi*: deceive.

In other instances, Hannan did not mark idiomatic expressions but gave them as citations for the verb. This is the case with the idiom *(ku)dya moto* (lit. eat fire, i.e. be harsh) which is found under the verb *(ku)dya* (eat), but without any indication that this sense is idiomatic. The entry is as follows:

- (6) **-dya** [H] KKoMZ v t & i Eat. 2. Defeat in a game [...] *Vanhurume vose vemana vanodyira padare*: all the males of the village eat at the com-

munal meeting place. [...] Anodya *moto*: he has a harsh manner of speaking (lit: he eats fire). [...]

With Hannan's method, it would not be easy for the user to find the idioms. Since the dictionary was targeted at users who wanted to learn Shona as a foreign language, the failure to label the fixed expressions accurately gives the impression that they do not exist or that they are not important.

In Dale (1981), idioms were also given as run-on entries. For example, the idioms *(ku)bata maoko* (to offer sympathy) and *(ku)bata kumeso* (to deceive) are both found under the verb *(ku)bata*, though Dale did not indicate that these are idiomatic usages but labels them as transitive verbs. The following entry shows how idioms are presented in Dale:

- (7) **-BATA** i & t (H) 1. [Bata maoko ake. Get hold of his hands.] vb *seize, catch, catch up* [...] 2 [Akabatwa namapurisa. *He was taken into custody by police.*] vb *arrest* [...] 4 **-b maoko** t [Ndinoda kundobata maoko kumusha. *I want to go and offer my condolences at home.*] vb *commiserate in bereavement* [...] 5 **-b chibharo** t vb *rape*; 6 **-b kumeso** [Akandibata kumeso ndikamupa mari. *He flattered me and I gave him the money.*] vb *deceive by flattery, flatter* [...]

There are no significant differences between Hannan's and Dale's treatment of proverbs and idioms. They both interchange the two methods identified by Čermák (2000: 490). With these methods, the user's search for the proverbs would be more laborious and cumbersome. Although seemingly straightforward, these methods are in fact circular and inconsistent because there is no rule for determining which is the so-called semantically heavy word but it is left to the individual lexicographer's decision. For example, in the case of *-chenga* 'keep close' in (1) and (3), the definition of the proverb *Chenga ose manhanga, hapana risina mhodzi* could be found either under the verb *-chenga* 'safeguard' or under the nouns *manhanga* 'pumpkins' or *mhodzi* 'seed'. This method is sufficient for the user who is acquainted with the full form of the proverb, but in cases where only a part of it is known, it will be very difficult if not impossible for the user to find it when it is treated as a run-on entry. The other weakness of this method is that information is more difficult to access because of the denseness of the entries into which different kinds of information are concentrated.

In *Duramazwi reChiShona* (1996), the predecessor to the DGS, some idioms were entered but not marked. They were given as run-on entries of the respective nouns or verbs from which they are derived. Thus, an idiom like *(ku)dya marasha* (lit. to eat fire, i.e. have a harsh temperament) (sense 5 in example (8)) is found under *(ku)dya* (eat).

- (8) **-dya** K it. 1 Kana *uchidya unenge uchiisa chikafu mumuromo, wotsenga uchimedza*. (*When you are eating you are putting food in your mouth, then*

you chew and swallow.) 2 Kana wadya vamwe pamakwikwi unenge wavakunda. (*When you have eaten others in a competition, you have defeated them.*) [...] 5 Kana munhu achinzi anodya mbwa, marasha kana gaka anenge achipenga zvikuru. (*If a person is said to eat a dog, embers or cucumbers, it means that they have a harsh temperament.*)

Entering idioms in this way would not have been easy in the DGS, because of the large number of idioms included. This method, as mentioned earlier, is also not user-friendly. That is why it was decided to give idioms lemma status in the DGS.

4. The treatment of fixed expressions in *Duramazwi Guru reChiShona*

4.1 Proverbs

Shona proverbs (*tsumo*, abbreviated *ts*) are of two types: short sayings and complete statements having a bipartite structure. Examples of short sayings are:

- (9) **Rudo ibofu.** *ts.* Iyi itsumo inoreva kuti kana munhu adzamirwa norudo haazombooni kuti pane chimwe chinhu chakaipa pamudikani wake. (*lit. Love is blind.*)
- (10) **Atsunya arwa.** *ts.* Iyi itsumo inoreva kuti munhu anoita kanhu kadikidiki pane chimwe chiiitiko chihombe anenge aitawo pake ipapo zvekuti vanhu havafaniri kumushora. (*lit. He who has pinched has fought, i.e. he who has done little compared to others should not be despised.*)
- (11) **Kunonoka huvhizhura.** *ts.* Iyi itsumo inoreva kuti dzimwe nguva kunonoka ndiko kunowanisa munhu zvinhu zvakanwanda pane kukasira. (*lit. To be late is a form of speed, i.e. at times being late in doing something can be rewarding.*)

Such short sayings as given in examples (9)–(11) do not pose serious challenges in lemmatisation, because they can be compared to compounds or collocations in terms of the space the lemma takes up. Fortune (1975–1976: 26) refers to this type of proverbs as "verbal sentences".

The larger number of proverbs which have a bipartite structure were entered in shortened form. The shortening of the proverbs was not done haphazardly but the first proposition of the proverb was entered as lemma. The full form of the proverb was then given after the first proposition, indicated with a bullet (•) as marker, followed by the explanation in brackets, as shown in examples (12) and (13) below:

- (12) **China manenji hachifambisi.** *ts* • *China manenji hachifambisi, chinomirira kuti mavara acho aonekwe.* (Nguva zhinji, kana paine chinhu chikuru

chinoda kuitika, kunotanga kuitika zvinhu zvidiki zvinoshamisa icho chisati chaitika.) (*lit. A tragedy or catastrophe is usually preceded by astonishing minor incidents.*)

- (13) **Chada mwoyo hachikoni.** ts • *Chada mwoyo hachikoni, nokuti mwoyo chiremba. [...]* • *Chada mwoyo hachikoni, mwoyo ndishe.* (Kana munhu aine chinhu chaanoda kuti ave nacho, anoita zvose zvaanogona kuti achiwane chero zvakaoma.) (*lit. A thing that the heart desires will be accomplished because the heart is the king, i.e. if a person yearns for something they will strive to achieve it.*)

Example (13) shows that all proverbs having the same beginning and the same meaning were included under one entry.

4.2 Idioms

Idioms in Shona (*madimikiva*, abbreviated *dimik*) take the form of verbal phrases, that is, the initial word in the idiom is often an inflected verb. The trend in previous dictionaries was therefore to enter the idiom under the verb controlling its meaning. In the DGS, however, idioms were given lemma status. Idioms in Shona all begin with the infinitive verb *ku-* 'to'. The verb stem was therefore hyphenated and entered in its alphabetical place as can be seen from the entries of the verbs *-bata* 'catch, hold' and *-dya* 'eat' and the various idioms derived from these verbs. Because of space constraints only a few examples of the general definitions of the verbs are given (see examples (14)(a)–(d) and (15)(a)–(d)). In the DGS, the verb *-bata* has 18 senses and there are 46 idioms beginning with it. The same applies to *-dya* which has only 4 senses, but forms the beginning of 42 idioms. If these idioms had all been treated as run-on entries under the verbs *-bata* and *-dya*, these lemmata would have been very congested.

- (14) (a) **-bata** K it. Kugunza noruoko kana kuisa muchanza. (*To touch or to hold in your hand.*) 1 Kubata kusimudza kana kutsigira chinhu nemaoko. (*To hold something is to lift it or to hold it in your hand.*) 2 Kubata kuti dzvi chinhu chawanga uchitandanisa (*To catch something you were chasing.*) [...] 10 Kubata kuona kana kunzwisisa zviri kutaurwa kana kuti zviri kuitika (*To understand or comprehend what is being said or what is going on.*) [...] 18 Kana sango richinzi rakabata rinenge rine miti yakawanda, mikuru, iri pamwe chete, uye richityisa kupinda. (*When a forest is said to be held together (i.e. to be dense) it means that it has a lot of big trees that make it scary to go into that forest.*)
- (b) **-bata chigaro** K KDD dimik. Kana munhu akabata chigaro anenge ane chinzvimbo choutungamiri. (*When someone is holding a seat they are in a position of leadership.*)

- (c) **-bata denga** K KD dimik. Kana munhu akafara zvekubata denga anenge afara zvakanyanya. (*If someone is happy to the point of touching the sky, it means they are very happy.*)
- (d) **-bata kumeso** K KDK dimik. Kana munhu achikubata kumeso anenge achinyepedzera kunge ari kuita zvakakunakira asi iye achikunyengetedza. (*If someone is said to be holding your face (i.e. deceiving you), they are pretending to be assisting you yet they will be deceiving you.*)
- (15) (a) **-dya** K it. Kutora chikafu uchimedza. (*To put food into your mouth and swallow it.*) 1 Kana uchidya unenge uchiisa chikafu mumuromo, wotsenga uchimedza. (*When you are eating, you are putting food in your mouth, then you chew and swallow.*) [...] 2 Kana wadya vamwe pamakwikwi unenge wavakunda. (*When you have eaten someone in a competition, you have beaten/defeated them.*) [...]
- (b) **-dya cheziya** K KDK dimik. Kudya cheziya kuzowana chokuzviraramisa nacho wamboita chokushandira zvakasimba. (*To eat something of sweat means that in order to achieve something you have to work hard for it.*)
- (c) **-dya marasha** [-dya maratya] K KDD dimik. Kudya moto. (*To eat fire.*)
- (d) **-dya moto** [-dya mwoto] K DK dimik. Kudya moto kupengera munhu uchimupotera zvikuru, washatirwa. (*To eat fire is to shout at someone when you are angry.*) [...]

4.3 Pithy sayings

Pithy sayings, categorised as *kachirevo* (abbreviated *kr*), were entered in the second section of the dictionary as part of the macrostructure. In form they resemble both proverbs and idioms. Some of them are two-word phrases such as (16), while others are complete statements such as (17) and (18).

- (16) **Charasika chayerera.** kr. Aka kachirevo kanotaura kuti kana chinhu chaitika chatopfuura, hamufaniri kuramba muchinetseka nezvacho. (*This is a pithy saying that says that what is done is done and people should not worry about what they can no longer change.*)
- (17) **musha mukadzi** kr. Kutu *musha mukadzi* kureva kuti, kuti murume ave nechiremerera, anofanira kuva nemudzimai anenge ari iye muchengeti wake anoona kuti basa repamba rafamba here. (*To say that a woman is what makes a home means that for a man to be respected socially, he has to have a wife to look after him and to take care of the home.*)
- (18) **imbwa hairangwe nekudimurwa muswe** kr. Ichi chirevo chinoreva kuti chero munhu akakanganisa zvakadii asi haapiwi chirango chinozomurwadza kweupenyu hwake hwose. (*This is a saying which means that if someone wrongs you do not give them a harsh penalty that will affect them for the rest of their life.*)

5. Conclusion

A dictionary should be compiled in such a way that it is user-friendly. The best way to achieve this is to ensure that the user finds what he/she is looking for. The most effective way to achieve this is to ensure that the information in the dictionary is easily accessible to the user. This article has shown that in lemmatising fixed expressions such as proverbs, idioms and pithy sayings, *Duramazwi Guru reChiShona* attempted to avoid circularity, making it easier for the user to retrieve the information. The article has furthermore indicated that taking the target users of the dictionary, who are mainly students, into consideration also influenced the inclusion of fixed expressions as lemmata.

Bibliography

Dictionaries

- Chimhundu, H. (Ed.).** 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.).** 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Dale, D.** 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gweru: Mambo Press.
- Hannan, M.** 1959. *Standard Shona Dictionary*. Salisbury: College Press.
- Hannan, M.** 1974. *Standard Shona Dictionary*. Second edition. Salisbury: Rhodesia Literature Bureau.
- Sinclair, J.M. et al. (Eds.).** 1995². *Collins COBUILD English Language Dictionary*. London: Harper-Collins.

Other sources

- Burkhanov, I.** 1998. *Lexicography: A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Čermák, F.** 2000. Combination, Collocation and Multi-Word Units. Heid, U. et al. (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th, 2000*: 489-495. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Chimhundu, H.** 1980. Shumo, Tsumo and Socialization. *Zambezia* 8(1): 37-51.
- Fortune, G.** 1975-1976. Form and Imagery in Shona Proverbs. *Zambezia* 4(2): 25-55.
- Gouws, R.H.** 2003. Types of Articles, their Structure and Different Types of Lemmata. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 34-43. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.
- Gouws, R.H. and D.J. Prinsloo.** 1997. Lemmatisation of Adjectives in Sepedi. *Lexikos* 7: 45-57.
- Gouws, R.H. and D.J. Prinsloo.** 2005. Left-expanded Article Structures in Bantu with Special Reference to IsiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.
- Kipfer, B.A.** 1984. *Workbook on Lexicography: A Course for Dictionary Users with a Glossary of English Lexicographical Terms*. Exeter Linguistic Studies 8. Exeter: University of Exeter.

- Lorentzen, H.** 1996. Lemmatization of Multi-word Lexical Units: In which Entry? Gellerstam, M. et al. (Eds.). 1996. *EURALEX '96 Proceedings: Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 415-421. Gothenburg: Department of Swedish, Gothenburg University.
- Moon, R.** 1992. 'There Is Reason in the Roasting of Eggs': A Consideration of Fixed Expressions in Native-speaker Dictionaries. Tommola, H. et al. (Eds.). 1992. *EURALEX '92 Proceedings. Papers Submitted to the 5th EURALEX International Congress, Tampere 1992*: 493-502. *Studia Translatologia* A.2. Tampere: Department of Translation Studies, University of Tampere.
- Svensén, B.** 1993. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Zgusta, L.** 1971. *Manual of Lexicography*. *Janua Linguarum. Series Maior* 39. Prague: Academia / The Hague/Paris: Mouton.

Some Deficiencies in the Mediostructure of *Isichazamazwi SesiNdebele*

Dion Nkomo, *African Languages Research Institute (ALRI), University of
Zimbabwe, Harare, Zimbabwe (deeouf@yahoo.co.uk)*

Abstract: In this article, a descriptive and critical analysis of the mediostructure of *Isichazamazwi SesiNdebele*, henceforth the ISN, is given. This analysis is made within the theoretical premises of the mediostructure and its functions. The aims of the article are to shed light on the following questions: What aspects of the mediostructure are used? Why are they used? How effectively are they used to achieve text compression/condensation and text cohesion? What improvements could be made to obtain greater user-friendliness? Such an analysis, it is hoped, will, in two ways, contribute to the development of lexicography in general and Ndebele lexicography in particular. Firstly, it is hoped to make users aware of the importance of some guidance aspects of the mediostructure to improve dictionary use. Secondly, it is hoped the pointing out of some deficiencies and possible improvements in the ISN mediostructure will be useful for the compilation of future Ndebele dictionaries.

Keywords: DICTIONARY, DICTIONARY STRUCTURE, MACROSTRUCTURE, MICROSTRUCTURE, MEDIOSTRUCTURE, CROSS-REFERENCING, ACCESS STRUCTURE, USER-FRIENDLINESS, FRONT MATTER, BACK MATTER

Opsomming: 'n Aantal gebreke in die mediostruktuur van *Isichazamazwi SesiNdebele*. In hierdie artikel word 'n beskrywende en kritiese ontleding van die mediostruktuur van *Isichazamazwi SesiNdebele*, voortaan die ISN, gegee. Die ontleding word gedoen binne die teoretiese gegewe van die mediostruktuur en sy funksies. Die doelstellings van die artikel is om lig te werp op die volgende vrae: Watter aspekte van die mediostuktuur word gebruik? Waarom word hulle gebruik? Hoe doeltreffend word hulle gebruik om teksverdigting/-verkorting en teks-samehang te bereik? Watter verbeteringe kan gemaak word om groter gebruikersvriendelikheid te verkry? So 'n ontleding, word gehoop, sal op twee maniere bydra tot die ontwikkeling van die leksikografie in die algemeen en van die Ndebeleleksikografie in die besonder. Eerstens word gehoop dat dit gebruikers bewus sal maak van die belangrikheid van 'n aantal leidingsaspekte van die mediostruktuur om woordeboekgebruik te verbeter. Tweedens word gehoop dat die uitwys van 'n aantal gebreke en moontlike verbeteringe in die ISN-mediostruktuur nuttig sal wees vir die samestelling van toekomstige Ndebelewoordeboeke.

Sleutelwoorde: WOORDEBOEK, WOORDEBOEKSTRUKTUUR, MAKROSTRUKTUUR, MIKROSTRUKTUUR, MEDIOSTRUKTUUR, KRUISVERWYSING, TOEGANGSTRUKTUUR, GE-BRUIKERSVRIENDELIKHEID, VOORWERK, AGTERWERK

1. Introduction

Dictionary structure is a case of convention with 'dictionary' having become associated with a certain structure peculiar to it. For instance, viewed superficially, the megastructure, i.e. "the overall framework of the component parts" of most dictionaries is similar, because they carry the 'front matter', the 'middle matter' and the 'back matter' texts (Hartmann and James 1998: 92). According to Kammerer and Wiegand (1998, referred to by Gouws 2004), a dictionary is viewed as a carrier of texts. The texts themselves convey information to the user. The information is arranged and treated systematically according to the conventional structure of a dictionary. At a macrostructural level, a certain method of arranging lexical items, be it alphabetising, niching or nesting, has to be used. Also necessary is a systematic provision and treatment of information categories of identical lexical items. These various information categories do not occur in this order and structure in their mother contexts.

A dictionary therefore is the result of taking pieces of knowledge from their larger contexts and arranging them in a conventional structure. This data distribution structure in dictionaries leads to decontextualisation of information. It is the lexicographer's role to make this information meaningful outside its context by following the conventions of dictionary structure. The metalexigraphers' role is then to make practicing lexicographers aware that the dictionary structure has a certain purpose. Metalexigraphers have emphasised that a dictionary is not merely a carrier of texts; it is also a carrier of information (Gouws 2004, Gouws and Prinsloo 1998, 2005), a container of knowledge (McArthur 1986) and, according to Wiegand (1988, referred to by Tarp 2000), a utility product with a genuine purpose. The texts should therefore convey useful information to the user and exhibit some form of network which helps the user obtain optimum information from the dictionary.

The lexicographer and the practice of dictionary making exist between two often opposing forces. On the one hand are the principles prerequisite for dictionary making and on the other hand is the society often unfamiliar with the principles. Playing a balancing role, the lexicographer has to convey information to this society within the confines of the dictionary type and size and the lexicographic methods and conventions in order to meet its reference needs and skills. Given such a scenario, no perfect balance is always struck in producing a dictionary of good quality meeting the user needs and skills. While the lexicographer has to serve the society through his practice, he has no license to violate the lexicographic methods and conventions at will in producing a usable dictionary for his society. The innovations should always be carried out within the principles of dictionary making, not disregarding them in a quest to achieve more user-friendliness.

In this article, the mediostucture is brought to the fore as an important convention central to lexicography as it serves both the maker and the user of the dictionary. It is viewed as an intersection of two practices of great disparity:

dictionary making and dictionary use. As a convention, it assists the lexicographer to gather information from different sources, compress and condense it in order to save space, and distribute it in the different access positions following a clearly spelt out macro- and microstructural system. The success of the ISN is analysed in this regard, also considering the accessibility of information.

In the analysis, the following questions are addressed: What aspects of the mediostructure are used? Why are they used? How effectively are they used to achieve text compression/condensation and text cohesion? What improvements could be made to enhance user-friendliness at the mediostructural level? Such an analysis is of particular importance to Ndebele lexicography as guidance for both the lexicographer and the user. The user is guided with the help of the mediostructure which should improve dictionary consultation. Critically viewing the mediostructure gives insights which lexicographers can consider in the production of future Ndebele dictionaries, particularly an *Advanced Ndebele Dictionary* (AND).

2. The Principal Functions of the Mediostructure in a Dictionary

This section does not intend to explore the mediostructure theoretically with a view to criticising and developing it. Only an overview based on theoretical and critical insights provided by Wiegand (2004), Gouws (2004), and Gouws and Prinsloo (1998, 2005) is given to create a formal base for the analysis of the ISN mediostructure. To arrive at an informed analysis, a clear description of the mediostructure and its functions is needed.

According to Gouws and Prinsloo (1998: 17), the mediostructure is the system of cross-referencing used to establish relations among different components of a dictionary. In simpler terms, it is a cross-reference structure by which is meant "the network of cross-references which allows compilers and users of a reference work to locate material spread over different component parts" (Hartmann and James 1998: 32).

In the introduction, it was mentioned that a dictionary is a carrier of texts of information. By following a conventional structure in its making, a dictionary might contain related information in different texts or components of texts. Although the different dictionary texts or components of texts may contain similar or related information on language, the information would always be treated differently to avoid repetition and redundancy. A dictionary with a clear function would have a way of connecting the texts, for they are not totally independent of each other. A good quality dictionary would have, among other features, a systematic and consistent way of establishing the relationship of the dictionary texts and their dependency on one another.

When the microstructural level is considered as an example, text compression and condensation through the use of symbols and similar methods and conventions are necessary to save space. Valuable as they are to any lexicographer, the methods and conventions need to be clearly accounted for and ex-

plained in the dictionary, usually in the front matter as guide to dictionary use. According to Gouws and Prinsloo (1998: 28), "cross-references from the front matter, especially from the user's guidelines to the central text are crucial to the user for successful or optimal retrieval of information". In this regard, they (Gouws and Prinsloo 1998: 34) add:

The front-matter texts should also include a systematic exposition of other language-specific characteristics and these have to be addressed from the central list by means of a well-developed mediostructural network.

The language-specific features should not be left unconsidered; if included, they should be explained in the front matter of the dictionary. For instance, if the front matter of a dictionary mentions orthographic problems, they should be treated consistently throughout the dictionary as explained in the front matter.

Similar principles may also be advanced with regard to the macrostructure. The popularly used alphabetical ordering presents words as if they occur as isolated units in a language, yet they are "socialized linguistic forms compiled from the speech-habits of a given speech-community" (Zgusta 1971: 17). In their socialisation, the words of a language are related as synonyms, antonyms, homonyms and so on. The dictionary structure, particularly the macrostructure, can obscure all such kinds of natural relations among the vocabulary items of a language. Wiegand (2004: 218) therefore says that "in order to uncover the connection, which is necessarily concealed by data distribution along the different access positions, a mediostructural network has to ensue". In such instances, the mediostructure would also help avoid repetition and redundancy thereby saving dictionary space. Re-establishing natural relations of linguistic and conceptual systems with a consistent use of cross-referencing devices produces a coherent and user-friendly dictionary.

However, if user-friendliness levels are to be increased, great care has to be taken in the creation of a mediostructure. This concerns the reference markers and their purported functions. Reference markers are typographic and non-typographic cross-reference devices which guide the user from a reference position to a reference address. These may be words like SEE, ALSO, GO TO, CHECK, COMPARE and many such ones in English, or symbols like an arrow pointing to another word which will be the reference address from the reference position. The communication of these needs to be explained in the front matter and their application to similar entries needs to be consistent with the explanation in the front matter. The lexicographer has to be reliable, i.e. if a user moves from one point to another as advised, the user should in fact find clear or more explicit information at that point.

Cumbersome circular referencing whereby the user is advised to find the meaning at another point, but, on arriving there, is again informed to look for the meaning at the first point, should be avoided. In fact, users resent being referred to another article to find information, when it takes more time than

planned to consult a dictionary. It becomes worse in cases of cross-referencing where the reference address is no better than the reference position in terms of clarity and informativeness.

Moreso, reference markers may increase the density of the microstructure. Great caution is necessary in the selection of the markers so as to avoid those that may obscure other important microstructural entries, thereby compromising the user-friendliness of a dictionary.

3.1 *Isichazamazwi SesiNdebele (the ISN)*

Not all of the details on the background of the ISN and its compilation are helpful in clarifying the argument in this article. Only those affecting the mediostructure are given. For more comprehensive details on the background of the ISN, the thesis of Hadebe (2002) may be consulted. Also, a number of articles such as Hadebe (2004) and Ncube (2005) give detailed background information on the dictionary and its compilation. Nkomo (2003) analysed its microstructure, showing that the microstructure and the mediostructure interact at some point.

The ISN is the first and so far the only monolingual dictionary in Ndebele. It is a medium-sized general-purpose dictionary targeted at Ndebele speakers in general, and students at schools, colleges and universities in particular. The dictionary is corpus-aided to account for some social and regional variants, and some loan- and controversial words in Ndebele, whose speakers are a heterogeneous group. As in all Nguni languages, the ISN compilers, as shown by Maphosa (1997, 1999), encountered some lemmatisation problems, especially with regard to nouns and verbs.

The mediostructure of the ISN, as will emerge shortly, was partly influenced by some of the factors mentioned above. Compiling the first monolingual dictionary, the ISN editors needed to be cautious when creating the mediostructure and combining different dictionary texts to achieve user-friendliness.

3.2 **The ISN Mediostructure**

This descriptive analysis of the ISN mediostructure is undertaken at different structural levels, namely the macrostructure, the microstructure, the access structure as well as the dictionary as a unit. Given the theoretical base overviewed in the previous section, the mediostructure is expected to be a space-saving and cohesive device which, above all, facilitates users' access to dictionary information.

At a macrostructural level, it generally seems to be every lexicographer's ambition to achieve a wide coverage of the vocabulary of the language. Some lexicographers, especially in the U.S.A., would inflate figures by counting the

word and its derivatives, eventually boasting of the number of entries contained in their dictionaries (Béjoint 2000: 43 and Landau 2001: 29). Although the ISN editors could have wished for a more extensive dictionary, its size (its main text) was restricted to a maximum of 545 pages (Hadebe 2002). Carefully formulated headword selection and lemmatisation procedures had to be followed in the compilation process. Hadebe (2002: 91) states, for instance:

(Only) nouns denoting singular were entered except in cases where nouns denoting plural were commonly used or where only the plural forms currently exist.

Entering nouns in both their singular and plural forms throughout the dictionary would have taken up much space, but would have added little in covering the Ndebele lexicon. In cases where both the singular and plural forms were lemmatised, repetition and redundancy had to be heeded. Through the mediostructure, such entries also had to be linked to indicate their singular/plural relationship.

Synonyms and variants too form part of the macrostructure of the ISN. The following principles guided headword selection and lemmatisation with regard to synonyms and variants for the ISN (according to the Style Manual for the ISN, quoted from Hadebe 2002: 237):

- (a) Any synonym should be entered as a separate headword.
- (b) Any variant should be entered as a separate word.

Unlike the principle for lemmatising nouns quoted above, headword selection and lemmatisation principles regarding synonyms and variants imply repetition of similar words, or words with the same meanings. For whatever reason, repetition is not a good practice in lexicography considering the importance of dictionary space. Where it is practised, it should be justified convincingly.

Briefly focusing the discussion on some types of macrostructural entries has shown the need for cross-referencing for reasons of space economy and establishing relationships. The sets identified include singular/plural forms, synonyms and variants. By looking at the method of cross-referencing among the identified sets, microstructural categories will inevitably be discussed as well. The discussion will be directed by the following examples and others in the appendix which lists articles having some irregularities with regard to the mediostructure.

- (1) **iliva** [iva] ... BONA *ameva*. (**thorn** ... SEE *thorns*.)
- (2) **amabhanga** ... 2 BONA *ibhanga*. (**banks** ... 2 SEE *bank*.)
- (3) **isenku** ... BONA *ithonga*. (**ghost** (S1) ... SEE S2.)
- (4) **ithonga** ... BONA *idlozi*. (**ancestor** (S1) ... SEE S2.)
- (5) **idorobho** ... BONA *idolobho*. (**town** (V1) ... SEE V2.)
- (6) **ijiphi** [ijibhi]... BONA *ijibhi*. (**jeep** (V1) ... SEE V2.)

Key: S1 – main entry
S2 – lesser synonym
V1 – main entry
V2 – unpopular variant

A brief glance at examples (1)–(6) and many similar ones on every page of the ISN is likely to establish the belief that cross-referencing is an effective space-saving device in the dictionary. The ISN editorial team itself, as Hadebe (2002: 101) rightfully states, is convinced of this:

The extensive use of cross-referencing enabled the lemmatization of up to twenty thousand entries in a six hundred page dictionary. It . . . saves space and avoids repetition of definitions.

While this is true, it may be realised that with greater caution in adhering to the principle behind cross-referencing, much more entries could have been conveniently added without making the dictionary bigger. There are numerous cases of repetition. For instance, some cross-referenced synonyms like **ikhehla/ixhegu** (old man), **abalози/idlozi** (ancestor(s)) and variants like **isingake>thiso/isingathekiso** (metaphor) in each case carry the same definitions. When it comes to defining, one article should carry the definition while it should be indicated as part of its microstructure that it has a relationship with the other which it defines. These examples run counter to the editorial policy of employing cross-referencing to save space and avoid repetition of definitions.

There is also no need to treat a lexical item, say a variant, as an unpopular form using the popular form as definition, but giving as part of its microstructure an indication that the popular form is its variant. Many variants and synonyms in the ISN have been treated as exemplified in (7):

- (7) **amadilikelana** [amadilikelane] bz 6. *Amadilikelana* ngamathumbu amancane aseduze lamalulu alobulongwe obulula.
amadilikelane [amadilikelana] bz 6. BONA *amadilikelana*.

If **amadilikelane** is a variant of **amadilikelana** (small intestines in animals like cattle), **amadilikelane** should carry a definition indicating **amadilikelana** as its variant, but it needs not also to be stated that the former is a variant of the latter.

Since it is given in the initial entry that **amadilikelane** is a variant of **amadilikelana**, it is also given that the reverse is true, making it redundant to indicate it in the latter article. With this treatment, the latter article would take up only one line. At least fifteen instances were identified in the first three pages of the dictionary, which, if this pattern is followed in the rest of the dictionary, would amount to a considerable number over three hundred pages. This seems to be more than an accidental mistake in the ISN mediostructural system. As such, the use of the mediostructure to save space is not exploited to the maximum, especially with regard to variants and synonyms.

'BONA' is the reference marker for all instances of cross-referencing in the ISN. As shown in the glosses of examples (1)–(6), it means 'SEE', which directs the user to the reference target for the definitions or more explicit information. This is satisfactory since the reference mark is a very simple word whose guidance cannot confuse any user.

What is likely to reduce user-friendliness of the mediostructural system of the dictionary is the impression given that there is only one form of relation between the cross-referenced sets, which is not true. Since these relationships have different ways of representation in the dictionary, this could also have been applied to their defining. Synonyms are represented by 'FAN' which represents 'amagama alengcazelo efanayo', meaning 'synonyms' in English. Variants are placed in square parentheses immediately after the headwords. Apart from the prefixes, there is no clear way of representing singular/plural nouns, but this is not likely to trouble Ndebele speakers.

Since the relations that necessitated cross-referencing in the ISN are different and deservedly differently represented, would it not have been also prudent for the editors to use different reference markers when cross-referencing them? There is a wide range of guide words used as reference markers in English dictionaries. Ndebele cannot be seen wanting in this regard. 'KHA' for 'KHANGELA' which is a near synonym to 'BONA' could have been used for one set (singular/plural nouns or variants) so that 'BONA' is not overburdened with different relations. Lesser synonyms could have been defined in the form of short sentences like "*Ikhehla lixhegu*" which would still be an implicit form of referring the user to the fully defined stronger synonym without taking up extra space. This method has partly been used in *Isichazamazwi SezoMculo* (ISM), the Ndebele musical terms dictionary, with the aim of achieving greater user-friendliness. If this proves to be the case, it should be considered in the compilation of the AND.

Apart from being useful to the dictionary-maker as a way of saving space, the mediostructure is an integral part of the access structure of the dictionary. By *access structure* is meant, according to Hartmann and James (1998: 2), "the component parts of the overall design of a reference work which allow the user to search for a particular item of information". Therefore, dictionary-makers should not concern themselves so much about the number of entries they can accommodate in a dictionary if they employ a certain method of text compression and/or cross-referencing which leaves their target users in the lurch. Wiegand (2004: 218) emphasises that "a mediostructural selection should result which depends decisively on the primary dictionary functions and consequently on the intended user relation".

With reference to Wiegand and other metalexigraphers who champion the user-perspective in dictionary making, it has been reiterated that a dictionary is a utility product. For instance, Béjoint (2000: 133) stresses that "the dictionary is an object that is used, almost an artifact". For any dictionary to be useful, its information has to be accessible at all its structural levels.

The ISN editors' use of a simple word as a reference marker is to be commended, although the limitation of its application has been shown. The cross-referencing does not overburden the microstructure with alien markers and other unnecessary conventions, which may make the user's inner search path difficult.

However, the main reason for the employment of cross-references is, as Svensén (1993: 199) notes, that "some parts of meaning units may be treated at different places while users may assume that the information has been exhausted within one entry". So, when the user finds a reference from one article to another, at that article, there should be given more information, which would solve the user's problem. In the case of the ISN, for instance, when the user is referred from **iliva** (thorn) to **ameva** (thorns) for its meaning, a definition giving the meaning of **iliva** certainly has to be provided. There are numerous cases in the ISN which do not meet this requirement. A few will be discussed to illustrate some of the problems the user is likely to encounter.

Following the mediostructure, the user may discover that in some cases there is no difference between cross-referenced entries in terms of information provision and treatment. For instance, both **idolobho** and **idorobho** (town) or **amachaphachapha** and **amabalabala** (spots) lack definitions, in both cases, the one being referred to the other. This circular referencing leaves the user stranded. In other cases, like **isambuzi** being referred to **ithoyilethi** (toilet), the entry to which the user is referred is not even lemmatised, thus creating an equally futile search-path.

There are some other cases of cross-referencing in the ISN eventually causing the user's search-path to be winding, if not misleading. One illustrative example is cross-referencing in the article **isenku** (ghost), which is defined using its synonym **ithonga** (ghost). Yet, there are two articles of **ithonga**, the other referring to 'ancestor'. This means that the user has to look at both articles. The first article does not even carry the definition; it is defined using **idlozi** (ancestor). For more examples, the search-path of headwords in the appendix may be followed.

So far the analysis of the ISN mediostructure has focused on the central list. It has already been emphasised that the mediostructure should eventually serve, after having made the information accessible to the user, as a cohesive device binding the dictionary texts together. The ISN as a carrier of texts contains three texts, namely the front matter, the central text and the back matter. The front matter of the ISN, 47 pages in length, is informative on the Ndebele language and the dictionary itself, giving a historical outline of the language and its speakers, a grammatical outline, an insight into the methods used at various stages of the compilation of the dictionary, the metalanguage list and the guide to the use of the dictionary. The ISN front matter, being the largest in Nguni dictionaries, was deliberately made extensive considering the fact that Ndebele has had no grammar book or monolingual dictionary prior to the ISN, which necessitated a comprehensive but concise outline of the structure of the

language (Hadebe 2002: 106-107). In this regard, the front matter is a well-advised addition by the editors.

By enumerating in the front matter what information the dictionary contains, how it is presented and how the user can best access it, the editors managed to produce a self-contained and perhaps a useful and user-friendly reference work. The macro-, micro-, medio- and access structures are all explained and illustrated to ensure that the user is guided in all the search-paths. It is unfortunate that dictionary users often do not read this very informative text. Hopefully this habit of ignoring the front matter text would diminish as Ndebele speakers acquire more reference skills with the development of a dictionary culture.

However, as an innovation some texts like the guide to dictionary use, the outline of headword selection, defining formats and others that directly relate to and impact on dictionary use could have been strategically inserted in the main text to form what is called 'middle matter' (Hartmann and James 1998: 94) so that the user is persuaded to read them whilst searching for words. For instance, texts like **Ukwethulwa Kwamagama** (Presentation of Entries) or **Ukwethulwa Kwengcazelo** (Presentation of Definitions) could be inserted at the beginning of letters or even on the page marked Aa-Zz where the real dictionary begins. At such points, users would be more likely to read these informative texts, which are in danger of remaining unutilised in the front matter.

With regard to the front matter, still more effort could have been made at the mediostuctural level to utilise the information it provides. For instance, while it is a good idea to give information on Ndebele grammar first in the front matter through the grammatical outline and the metalanguage list and then in the main text as defined entries, the texts containing the information are not linked. Since space economy in the main texts renders definitions brief, it could have been possible to refer the user at the end of the definitions of grammatical terms to the grammatical outline for more detail. This in a way would make the user aware of the useful information contained in the front matter. Accessibility to this information would be improved, with the prospect even of improving both reference skills and dictionary culture.

The application of some crucial principles in the dictionary is not consistent with the information in the front matter, giving details about what the dictionary contains and how it may be accessed. Some irregularities have already been indicated in certain aspects of cross-referencing which might impede the user's consultation of the dictionary. There also seem to be other irregularities indicating a lack of unity between the different texts of the dictionary. One example is the inconsistent adherence to the Ndebele alphabet presented on page xxxi when spelling some country names in the back matter, for instance the Ndebele equivalents for Liberia, the Ivory Coast, Kenya, Nigeria and notably Swaziland. Even some loan-words in the main text exhibit sound clusters which are not part of the list given on page xxxii. Not everything presented in the introduction and guide to the ISN is followed consistently, which might seriously effect the unity of the dictionary.

4. Conclusion

This article has explored the mediostuctural aspects in the ISN, with particular focus on its weaker and impeding points, against the background of the principal functions served by the mediostucture in a dictionary as previously explained and demonstrated by Wiegand (2004), Gouws (2004), and Gouws and Prinsloo (1998, 2005) among others. Suggestions have been made how more positive results could have been achieved. It is hoped that for any form of revision of the ISN, the compilation of the AND and other future dictionaries in Ndebele, these suggestions would be considered. Lemmatisation of multi-lexical units like idiomatic expressions in the AND and other multi-lexical units in LSP dictionaries is likely to present more challenges at the mediostuctural level.

Finally, it needs to be emphasised that focusing in the article on the deficiencies in the mediostucture of the ISN does not mean that the dictionary is without merit. Actually, its vast contribution to the development of the Ndebele language and lexicography cannot be adequately quantified. Critical reviews such as this are part of this development. Research on and studies of the language and its use will certainly educate and uplift its speakers as the dictionary gives an endangered language extra strength to survive and develop.

References

- Béjoint, H.** 2000. *Modern Lexicography*. Oxford: Oxford University Press.
- Gouws, R.H.** 2004. Outer Texts in Bilingual Dictionaries. *Lexikos* 14: 67-88.
- Gouws, R.H. and D.J. Prinsloo.** 1998. Cross-Referencing as a Lexicographic Device. *Lexikos* 8: 17-36.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice in South African Lexicography*. Stellenbosch: SUN PReSS.
- Hadebe, S.** 2002. *The Standardisation of the Ndebele Language through Dictionary-Making*. D.Phil. Thesis. Harare/Oslo: University of Zimbabwe/University of Oslo.
- Hadebe, S.** 2004. Improving Dictionary Skills in Ndebele. *Lexikos* 14: 89-104.
- Hadebe, S. (Ed.).** 2001. *Isichazamazwi SesiNdebele*. Harare: College Press.
- Hartmann, R.K.K. and G. James.** 1998. *Dictionary of Lexicography*. London: Routledge.
- Kammerer, M. and H.E. Wiegand.** 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 14: 224-238.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.
- Maphosa, M.** 1997. *The Morphological Structure of the Noun and its Implications on the Ordering of Entries in Ndebele Dictionaries*. Unpublished B.A. Honours Dissertation. Harare: University of Zimbabwe.
- Maphosa, M.** 1999. *The Morphological Structure of the Verb in Ndebele and its Implications on Headword and Sense Selection*. Unpublished M.A. Dissertation. Harare: University of Zimbabwe.
- McArthur, T.** 1986. *Worlds of Reference*. Cambridge: Cambridge University Press.

- Ncube, C.** 2005. Language Development or Language Corruption? The Case of Loan-words in *Isichazamazwi SesiNdebele*. *Lexikos* 15: 384-395.
- Nkomo, D.** 2003. *The Microstructure and the Accessibility of Dictionary Information: An Analysis of Isichazamazwi SesiNdebele*. Unpublished B.A. Honours Dissertation. Harare: University of Zimbabwe.
- Nkomo, D. and N. Moyo (Eds.)**. 2006. *Isichazamazwi SezoMculo*. Gweru: Mambo Press.
- Svensén, B.** 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford: Oxford University Press.
- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Wiegand, H.E.** 1988. Was eigentlich ist Fachlexicographie? Mit Hinweisen zum Verhältnis von sprachlichem und enzyklopädischem Wissen. Munske, H.H. et al. (Eds.). *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern*: 729-790. Berlin/New York: Walter de Gruyter.
- Wiegand, H.E.** 2004. Reflections on the Mediostructure in Special-Field Dictionaries. Also According to the Example of the *Dictionary for Lexicography and Dictionary Research*. *Lexikos* 14: 195-221.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague: Mouton and Company.

Appendix

This is a list of some of the ISN headwords which have been found to be poorly cross-referenced. This means is that cross-referencing has not been exploited to the maximum to save space, because of repetition of synonyms or variants and definitions in a way not improving accessibility to information. In some cases, cross-referencing is ineffective in guiding the user, because of the absence of the reference address or the definition at the reference address, or because of further cross-referencing making the search-path longer. In other cases, cross-referencing is not clear enough to establish the intended relations between the reference position and reference address. It should be noted that the list only contains cases of poor cross-referencing up to page 12 of the dictionary.

ababusi	amagcikwane	amalwa ecatsha
abadala	amagciyane	amamkhwezikhwezi
abalozi	amahabula	amanangananga
abatshakazi	amahalawubho	amancoko
akunani	amahele	amandambi
alubana	amahemuhemu	amandiki
amabethe	amahlafuhlafu	amangebeza
amachaphachapha	amahlakomuzi	amangqanga
amachaphazi	amahlala	amangqolo
amacilikwane	amahlebezi	amanhlahkomuzi
amacokamo	amahlungelo	amanikiniki
amacubane	amajinkojinko	amanqe
amadabhuli	amakaka	amanqonjwana
amadakadaka	amakhafitha	amantanga
amadilikelane	amakhekheba	amantelo
amadunuselane	amakhosi	amanxa
amafa	amakhwakhwa	amanxe
amafelamlonyeni	amakhwengukhwengu	amapheluphelu
amafinyela	amalabi	amaphico
amafunyafunya	amalombolombo	amaphivane
amagaka	amalungiselo	

The Impact of Lexicographical Work on Language Use: The Case of Shona Monolingual Dictionaries in Zimbabwe*

S. Nyota, *Department of Curriculum Studies (shumirainyota2@yahoo.com)*
and J. Mapara, *Department of Languages, Literature and Music,*
(jacomapara@yahoo.co.uk),
Masvingo State University, Masvingo, Zimbabwe

Abstract: This article examines the impact of current lexicographical work in Zimbabwe on some sectors of Shona language use, namely education, media, medicine and local government. It looks specifically at Shona monolingual lexicographical projects completed by the African Languages Research Institute (ALRI), successor to the African Languages Lexical Project (ALLEX). It analyses how users of Shona in these particular sectors are responding to the different lexicographical products published by ALRI. The article maintains that Shona monolingual lexicography has resulted in language raising and awareness. It has also led to term creation and has contributed towards standardisation of the language. Shona has furthermore gained the abstractive power it needs to explain its own and other concepts. All these have caused diglossia leakage from Low (L) Shona to High (H) Shona in some areas of Shona language usage. The overall effect is that Shona is now used in some formal sectors such as the above-mentioned ones which previously were the preserve of English in Zimbabwe.

Keywords: MONOLINGUAL LEXICOGRAPHY, GENERAL DICTIONARIES, SPECIALISED DICTIONARIES, STANDARDISATION, LANGUAGE RAISING, LANGUAGE AWARENESS, LANGUAGE USE, SLCA, ALLEX, ALRI

Opsomming: Die invloed van leksikografiese werk op taalgebruik: Die geval van Sjona- eentalige woordeboeke in Zimbabwe. Hierdie artikel ondersoek die uitwerking van die huidige leksikografiese werk in Zimbabwe op sekere sektore van Sjonataalgebruik, naamlik opvoeding, media, geneeskunde en plaaslike regering. Dit kyk spesifiek na Sjona-eentalige leksikografiese projekte wat deur die African Languages Research Institute (ALRI), opvolger van die African Languages Lexical Projects (ALLEX), voltooi is. Dit ontlee hoe gebruikers van Sjona in hierdie besondere sektore reageer op die verskillende leksikografiese produkte wat ALRI gepubliseer het. Die artikel voer aan dat Sjona- eentalige leksikografie taalverhoging en

* This article was presented as a paper at the Eleventh International Conference of the African Association for Lexicography organised by the Tshivenda National Lexicographic Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 3-7 July 2006.

-bewustheid tot gevolg gehad het. Dit het ook gelei tot termskepping en het bygedra tot die standaardisering van die taal. Sjona het verder die abstraherende vermoë verkry wat dit nodig het om sy eie en ander konsepte te verduidelik. Dit alles het diglossielekkasie van Lae (L) Sjona na Hoë (H) Sjona op sekere gebiede van Sjonataalgebruik veroorsaak. Die algehele uitwerking is dat Sjona tans gebruik word in 'n aantal formele sektore soos die bogenoemdes wat vroeër die alleengebied van Engels in Zimbabwe was.

Sleutelwoorde: EENTALIGE LEKSIKOGRAFIE, ALGEMENE WOORDEBOEKE, GESPE-SIALISEERDE WOORDEBOEKE, STANDAARDISERING, TAALVERHOOGING, TAALBEWUSTHEID, TAALGEBRUIK, SLCA, ALLEX, ALRI

Introduction

In addition to compiling dictionaries, the African Languages Research Institute (ALRI) also carries out language research and documentation and for this it has full-time data entry operators. However, this article deals exclusively with ALRI's lexicographical work in Shona. From its inception in 1992, the African Languages Lexical (ALLEX) Project, now ALRI, has published six Shona and Ndebele monolingual dictionaries. The Shona dictionaries are *Duramazwi reChiShona* (Shona Dictionary) (1996), *Duramazwi Guru reChiShona* (Advanced Shona Dictionary) (2001) *Duramazwi reUrapu neUtano* (Dictionary of Shona Biomedical Terms) (2004) and *Duramazwi reMimhanzi* (Dictionary of Shona Musical Terms) (2005). For the largest part, the information in these publications is in Shona, but there are English glossaries. The publications in Ndebele are: *Isichazamazwi SesiNdebele* (Ndebele Dictionary) (2001) and *Isichazamazwi SezoMculo* (Dictionary of Ndebele Musical Terms) (2006). The following are ALRI's forthcoming Shona dictionaries: *Duramazwi reDudziramutauro neUvaranomwe* (Dictionary of Shona Linguistic and Literary Terms) and *Duramazwi reVana* (Shona Children's Dictionary).

In a public lecture delivered as part of the University of Zimbabwe's Golden Jubilee Celebrations, Chimhundu (2005a) assesses the impact of lexicography on the raising of the languages in Zimbabwe with particular reference to Shona. His conclusions are that lexicography has had a tremendous impact on the raising of Shona as language. This article assesses the impact of monolingual lexicographical work on Shona language use. Monolingual lexicography is undoubtedly yielding positive results as far as the use of Shona is concerned.

Language use refers to all areas in which a language functions. Normally a language is assigned duties through status planning which is a type of language policy. However, as noted by Chimhundu (1997), there is no language policy framework in Zimbabwe (Roy-Campbell and Gwete 1997). The nearest Zimbabwe came to a language policy is through the 1987 Education Act Chapter 55 (Amended 1991; Part XI, paragraph 55: 225-226). This Act gives English official status while Shona and Ndebele are accorded national status. It rein-

forces the diglossic relationship that exists between Shona and English in Zimbabwe. The term *diglossia* was first coined by Ferguson (1959) to describe language situations where a society employs two varieties of the same language "sufficiently distinct for lay people to call them separate languages" (Hudson 1996: 49). Ferguson calls these varieties H(igh) and L(ow). In a diglossic situation, they are used in a subset of mutually exclusive functions (Ferguson 1959, Fishman 1972). On the one hand, the high variety is associated with the superior functions of language namely parliament, news, announcements and other forms of journalistic writing, editorials in newspapers, academic, intellectual and formal discourse, court, legislative bodies, invitation cards, notices and advertisements. On the other hand, the low variety is used in such informal contexts as conversation at home, folk entertainment, addressing people at political gatherings, especially in the rural areas, and during parliamentary, senate and presidential elections, and instructions to workers in low prestige jobs.

The notion of diglossia as formulated by Ferguson was not extended to bilingual communities. His four case studies namely Egypt, Greece, Haiti and Switzerland concerned monolingual situations. Fishman (1972) proposed that Ferguson's 1959 model of diglossia be extended to bilingual situations. Whereas Ferguson (1959) is concerned with language varieties, Fishman (1972) relaxes this restriction and applies the concept diglossia to communities where two or more languages occur side by side. Fishman's (1972: 92) distinction of varieties is "along the lines of H(igh) languages on the one hand and L(ow) languages on the other". This is a reasonable extension of diglossia, especially in communities such as the one in which Shona found itself in the old Rhodesia where the colonial power enforced its own language as the language for formal purposes, thereby reducing the status of the indigenous languages. Styled along Fishman's (1972) notion of diglossia, this relationship in Zimbabwe is such that English has enjoyed high status (H), while Shona (together with Ndebele) remains at low status (L) (Chimhundu 2006, Veit-Wild 2006, Mashiri 2002).

ALRI's Shona monolingual lexicographical work is therefore taking place in a diglossic situation where there is no clear language policy, where there is not a linguistic normative body (Chimhundu 2006: 11), and where the Education Act serving as language policy relegates the Shona language. Despite this uneven situation, standardisation of the Shona language is taking place, as Chimhundu (2006) confirms when he states:

The standardisation of the nominal national languages is happening without official policy planning or policy framework.

Historical overview of lexicography and standardisation of Shona

Lexicography is one of the pointers to standardisation of a language. The others are orthography and grammar. When Doke (1931) worked on the codification of the Shona language, he recommended the preparation of a comprehensive

grammar and the compilation of dictionaries as a guide to literary work. Doke believed that carrying out such recommendations and implementing them would ultimately result "in the development of Shona as a great literary language" (cited in Chimhundu 2005a: 29). To this end, Chimhundu (2005a: 1) elaborates:

Since lexicography is the meeting point of all the disciplines that have a bearing on language, success in raising the status of a language and in diversifying its functions to cope with development and modernisation largely depends on advances made in lexicography and terminology in that particular language.

In the case of Shona, ensuing from Doke's recommendations, three bilingual dictionaries followed his 1931 orthography. These are Barnes (1932), Biehler (1950) and Wild (1953). Several other bilingual dictionaries and a grammar handbook were also produced after the revised orthography of 1955, for example, Hannan (1959, revised edition 1974), Dale (1975, revised edition 1981) and Fortune (1957, revised edition 1967). These developments in Shona lexicography coupled with developments in Shona orthography and grammar had an impact on the use of Shona. For instance, the first Shona novel, *Feso* by S.M. Mutswairo, was published in 1956, followed by B.T.G. Chidzero's *Nzvengamutsvairo* in 1957. The first poem was also written in 1956, H.W. Chitepo's epic *Soko Risina Musoro* (A Tale without a Head). The works of Fortune, Hannan and others impacted positively on the development of Shona fiction and poetry as well as the teaching of Shona as a subject and as a language. In the following years, other novels and poetry anthologies succeeded these first efforts by Mutswairo, Chidzero and Chitepo. Shona became examinable at Ordinary Level in 1958 and at Advanced Level in 1974 (Tsodzo 1992). However, these earlier developments in Shona grammar and lexicography were not immediately followed by an interest in the use of Shona in society in the same way as seems to be the case with the dictionaries resulting from Shona monolingual lexicographical work.

Recent developments in the field of Shona lexicography are those in which ALLEX/ALRI is involved, with a deliberate concentration on Shona monolingual dictionaries. Between 1992 and 1995, while work was being done on its first monolingual dictionary which was published in 1996, ALLEX carried out extensive research in all Shona-speaking areas in Zimbabwe. These research activities involved the grassroots, especially schools and colleges as well as university students with some of them participating as field research assistants. In this way, awareness of and improved attitudes towards Shona were created, boosting the participants' confidence in using Shona. The impact was great, because the publication of the first Shona monolingual dictionary was soon followed by the publication of grammar textbooks for Shona written in Shona. These are Mashiri and Warinda's "A" level textbook (1999) as well as Nyota's "O" level textbook (1999). At an "A" Level marking session in Harare in 2001 some markers who had been impressed not only by the work of Mashiri and Warinda (1999) and Nyota (1999), but also by *Duramazwi reChiShona* (1996) and

Duramazwi Guru reChiShona (2001) mooted the idea of a Shona language association. This idea caught on and led to the formation of the Shona Language and Cultural Association (SLCA) at an initial meeting held at the Harare Girls' High School in 2002 and a subsequent meeting which took place at the Midlands State University in 2003. Today SLCA has become a strong partner of ALRI in Shona language research and documentation activities.

It is important to note that not only Mashiri and Warinda but also Nyota were motivated and aided by ALRI's monolingual dictionary of 1996. Their publications in turn made itself felt among teachers and students of Shona and other interested parties. In conjunction with SLCA, ALRI works tirelessly, carrying out campaigns to raise awareness of the part that Shona could play in Zimbabwe. ALRI and SLCA have regularly held joint meetings and have embarked on joint Shona projects such as an advanced analysis text of Shona literature ready for publication by Mambo Press. Other ongoing joint projects by ALRI and SLCA include a Shona phonetics and phonology text and a new and standardised Advanced Level Shona grammar text. There is reason to believe that the ALRI-SLCA campaigns are effective, because they involve Shona language users and promoters such as publishers, teachers from primary school to tertiary levels, the relevant Government ministries, the National Arts Council of Zimbabwe (NACZ), the Zimbabwe Schools Examination Council (ZIMSEC) and students. Consequently most people are now aware of the positive role Shona could play in its speech community. As a result of ALRI's monolingual dictionaries and the campaigns by both ALRI and SLCA, many areas such as the above-mentioned are increasingly using Shona where previously English was employed. Also following the ALRI-SLCA collaboration, positive responses have been noted as far as the use of Shona is concerned. For example, after the publication of *Duramazwi reChiShona* (1996) and *Duramazwi Guru reChiShona* (2001), some medical practitioners headed by Professor N.Z. Nyazema of the University of Zimbabwe's Medical School requested ALRI to produce a Shona medical dictionary. This resulted in the publication of the specialised Shona medical dictionary *Duramazwi reUtano neUrapu* (2004). This first dictionary of biomedical terms has in turn sparked the interest of more specialists and other practising medical personnel. A good example is Dr. F. Madzimbamuto, who has been motivated to join ALRI for the compilation of a larger and specialised biomedical dictionary. His contribution so far includes *Muzongozo neHungunyoyo* (The Vertebral Canal and its Contents) (2006) which gives such clear definitions that lay people can understand them, this in spite of the fact that Shona was once considered a language without terms relating to technical and scientific fields such as medicine. Even medical personnel themselves now see the need to simplify the medical jargon for the benefit of their patients, who are largely Shona-speaking and often elderly. This is a welcome development, because patients often comment on medical doctors' use of unfamiliar English words. As a result grown-up children have to accompany their elderly parents to doctors in order to translate the doctors' language and terminology for them. In sixteen pages that include diagrams showing parts of

the vertebral canal and its contents, Madzimbamuto has succeeded in clarifying a body part that Mpofu et al. (2004: 41) originally defined in twenty-one words.

The impact of Shona monolingual lexicography cannot be overemphasised. Shona language communities have benefited greatly from this initiative of ALRI. Firstly, the dictionaries created the much-needed terminology in the teaching of Shona as a subject at different educational levels, that is, from grade one to tertiary levels. As noted earlier, this field is complemented by ALRI's prioritisation of and shift to specialised or terminological dictionaries in different domains of Shona language use, for example, medicine, language and literature, and music. This shows the practical approach of ALRI's research, as Chimhundu (2005a: 1) confirms when he says:

The primary motivation of the research programme has always been need filling. As a way of responding to the most urgent practical needs of the language communities ...

Secondly, the monolingual Shona dictionaries can also serve as a way of empowering the indigenous language communities in post-colonial Zimbabwe. The indigenous languages were downgraded during the colonial era, when English was prioritised. This was a time of strict diglossia with English as H and Shona as L. Monolingual dictionaries could be said to have effected a reversal of the diglossic relationship between Shona and English in some areas of language use in Zimbabwe, resulting in what Fasold (1984: 44) refers to as diglossia leakage. Fishman (1972) describes this occurrence as bilingualism without diglossia. This happens when a L(ow) language assumes the function formally reserved for the H(igh) language, as a result of the standardisation of the L language.

Chimhundu (2005a: 2) notes that monolingual lexicography has a particularly powerful impact during the language raising process:

The monolingual dictionary is not only standard setting, but it also gives a language the abstractive power that it needs to describe itself.

Following the ALRI-SLCA collaboration, there seems to be a growing interest in and general consciousness of the use of Shona in areas, which were originally dominated by English. Several other fields, including education, electronic and print media and national and local government, have also shown the same interest as noted in the domain of medicine.

Areas showing the impact of lexicography

Education

One area exhibiting a close link with Shona lexicography is that of education. Prior to the publication of monolingual Shona dictionaries, some elements of Shona language and literature were taught in English, the argument being that

these were too complex for Shona to handle. Research by Dube (1995) on the diglossic relationship between English and Shona revealed that her respondents claimed English to be more expressive than Shona and Shona to lack grammatical, linguistic and literary terms. As such the respondents preferred English to Shona as medium of instruction for Shona grammar, linguistics and literature. These respondents may have revealed negative attitudes towards Shona, but their responses also indicated the underdevelopment of Shona. Shona then indeed lacked the technical terms for grammar, linguistics and literature as well as for other subject areas.

The result was that Shona grammar was taught in English, both at secondary and tertiary levels. For examination purposes at secondary school level, there was the choice of answering questions in either Shona or English. The Shona option was an empty formality because the technical terms for the language were all in English. Even examiners' marking schemes were in English. At that time, all teachers' colleges as well as the University of Zimbabwe (UZ) taught all Shona courses in English. This practice continued even after independence, especially at the UZ's Department of African Languages and Literature. This was despite the existence of the Education Act of 1987 (Amended 1991), Chapter 55, Number 3, which states:

From the fourth grade, English shall be the medium of instruction; provided that Shona or Ndebele shall be taught as subjects on an equal time allocation basis as the English language.

The situation seems to have changed with the publication of the Shona monolingual dictionaries compiled by ALRI, which inspired Mashiri and Warinda (1999) and Nyota (1999) to write Shona grammar textbooks in Shona. This was a result of the influence of the publication of the first monolingual dictionary as well as interest generated during the extensive data-gathering exercise carried out by ALLEX prior to its publication. Consequently, at secondary school level, Shona grammar, language and literature are now being taught and examined in Shona (Secretary's Circular Number 1 of 2002). The underdevelopment of Shona is no longer an issue in this domain since lexicographical work at ALRI has created and standardised the technical terms, the lack of which was an impediment. Even ZIMSEC is now working in close conjunction with ALRI in this regard.

Another pointer to the impact of Shona lexicography on education is the publication of grammatical works. After standardising Shona orthography, one of Doke's recommendations was, as noted earlier, the writing of a comprehensive grammar as a guide to Shona literary work. George Fortune is someone who followed this consignment. His works include *An Analytical Grammar of Shona* (1955), *Elements of Shona* (1957), and *Shona Grammatical Constructions* (in two parts, 1984 and 1985). Other works are those by Chigidi (1988) and Chakamba et al. (1987).

These grammarians analysed Shona grammatical constructions using Eng-

lish. They confirmed Dube's (1995) findings that pupils and students of Shona grammar and linguistics believed the grammatical and linguistic concepts were too complex for Shona to handle.

However, following the publication of the first Shona monolingual dictionary, this belief and practice proved to be only a myth, because Shona grammatical terms were created and Shona was used to explain itself as shown in Mashiri and Warinda (*Dudziramutauro reChiShona 'A' Level*, 1999) and Nyota (*Dudziramutauro reChiShona 'O' Level*, 1999). Works on Shona orthography were also written in Shona. Among them are Magwa's *Manyorerwo eChiShona* (1999) and Chimhundu's *Nhaka yaDoke* (2005b). These developments point towards the growth and expansion of Shona. An author such as Chigidi (1988) has been asked by his publishers to rewrite the English section on Shona grammar in his book in Shona, ensuring that it is in accordance with the terms standardised by the Shona monolingual dictionary and the forthcoming dictionary of Shona linguistic and literary terms.

The electronic and print media

There is also evidence that Shona monolingual lexicography has had some effect on both the electronic and print media. Chimhundu (2006: 10) observes that in 1990 the percentage usage of the majority language, Shona, by the Zimbabwe Broadcasting Corporation (ZBC) was 55% on the African languages station, Radio Zimbabwe (then called Radio 2), 40% on the educational channel, National FM (then Radio 4), 1.5% on the almost entirely music station, Power FM (then Radio 3), and 1% on the general and predominantly English language station, Spot FM (then Radio 1). Around that time, television programmes in Shona and Ndebele were very few. Only 1.5% of motion pictures was in Shona. The circulation of English language newspapers such as *The Herald* was much larger than that of the only indigenous language newspaper *Kwayedza/Umthunywa*, catering for both Shona and Ndebele.

The language composition in the print media has since changed greatly with the local languages Shona and Ndebele given much more space. The newspaper *Kwayedza/Umthunywa* has since been separated into two newspapers, namely *Kwayedza* (for Shona) and *Umthunywa* (for Ndebele). The requirement that all broadcasting on radio and television should have 75% local content contributed to a dramatic increase in airtime for local languages, particularly Shona and Ndebele. Chimhundu (2006: 11) noted the benefits of this shift as follows: improved attitudes towards indigenous languages, and a greater use of local languages in public life and in business, especially advertising. Participants now also join freely and confidently in radio and television programmes using local languages. The effect of the 75% local content policy on the use of indigenous languages on radio and television must be acknowledged, but the contribution of the ALRI-SLCA alliance's awareness campaigns and the standardisation of Shona, partly as a result of monolingual lexicogra-

phy, should not be underestimated. Shona and Ndebele which are allocated much more space and time over the other indigenous languages are the most developed, with Shona having an edge over Ndebele even in airtime slots and standardisation. If the increase had nothing to do with the development of Shona, then the programmes would show many technical and terminological variations, inconsistencies and errors. However, this is not the case.

A common error is the use of the small letter "v" in the honorific prefix *Va* when it is used with people's names and surnames, e.g. *vaChagumachii* and *vaTodyei* instead of *VaChagumachii* and *VaTodyei* respectively. The error is caused by confusing the class 2 prefix *va-* which refers to number with *Va-* which is a form of address to show respect. It can be translated into English as "Mr" or "Mrs", depending on the gender of the addressee.

Word-division errors also occur frequently in Shona texts found in the media, for example in Shona advertisements in both the print and electronic media. Some advertisements on bill boards for Toyota Zimbabwe read *Toyota i mota ye vanhu* instead of *Toyota imota yevanhu* (Toyota is the people's car). Another example is the Zimbabwean television soap serial called *Tiriparwendo*, which should be written *Tiri parwendo* (We are on a journey). These errors cannot be blamed on lexicography. Shona word-division problems can be backdated to the early missionaries who attempted to reproduce Shona in writing. The orthographies of the early missionaries who, from around 1900, embarked on writing unified Shona, were divergent in word-division, which were disjunctive, "guided mainly by equivalent translations in English and thus splitting Shona words unnecessarily" (cited in Chimhundu 2005c: 24). Doke (1931) recommended a conjunctive way of writing, which he considered appropriate for an agglutinative morphological typology such as that of Shona. This was a simple rule. However, the then Shona Language Committee made Shona word-division rules much more elaborate and complex (Fortune 1972). This problem has persisted to this day as evidenced by the numerous word-division errors even in official documents. Dube (1995) who examined the problem concluded that it continues to exist, because, firstly, the rules are not taught at school and, secondly, they are too many and too difficult to commit to memory.

The other area, which shows an increased interest in using Shona, is that of advertising. In a strict diglossic situation, as noted earlier, advertising is the preserve of the H language or language variety. This situation used to prevail in Zimbabwe. Pongweni (1983), who analysed the language of advertising used only one Shona advertisement, as compared to the numerous English ones. Mutangadura (2001) had two Shona advertisements against five in English. The situation has now changed. The media, both print and electronic, is inundated with Shona advertisements, as can be seen when reading newspapers, watching television, listening to the radio or passing billboards. The impact of ALRI's work on the media is the result of the campaigns embarked on by the SLCA, which, in the days prior to its launch, invited all stakeholders, including those from the Ministry of Information.

Both the indigenous and multinational corporations show an increased interest in Shona advertisements. Social service providers and political propagandists also use Shona advertisements. Agricultural, communication, insurance and construction companies are some of the sectors that have numerous and effective Shona advertisements. An analysis of the advertisements shows that they attract attention. They arouse interest, stimulate desire and create conviction, drawing positive responses and reactions from their audiences. In other words, apart from the word-division problem, Shona advertisements compete favourably with those in English. The impact of these advertisements relies on the standardised terminology found in the monolingual dictionaries. This has helped them to reach out to the bulk of the population, since Shonas constitute more than 75% of Zimbabweans.

National and local governments

The other impact of monolingual dictionaries is noticeable from the way some local municipality councils now communicate with residents by distributing their notices and circulars in Shona. An example is the municipality of Masvingo that informs residents in Shona about issues relating to water use and sewerage pipes. The city of Harare also warns residents in Shona not to amass and dump rubbish in certain areas. The way the notices and circulars are composed clearly shows that the writers have benefited from the orthography and terminology resulting from ALRI/ALLEX. This becomes evident from the minimum errors these now contain, unlike in the past when mistakes were a common feature of the few notices and circulars that were distributed in Shona.

While the lexicographical work undertaken by ALRI has had a major impact on these critical sectors of Zimbabwean society, there are some areas where the use of Shona, despite its development as a standardised language, is being resisted. Some denominations and especially the courts are examples. Like these denominations, the courts rely on translators, even where the plaintiff, the defendant, the prosecutor and the judge or magistrate are Shona-speaking. Hopefully the courts will also realise the importance of empowering those who attend the courts for various reasons. This can be done through the use of the language they are most comfortable with and understand best.

Conclusion

This article has shown that monolingual lexicography has had a great and positive impact on Shona-speaking communities in Zimbabwe. Its influence has been realised in areas like the media, education, medicine and local government. Initially lexicography has aided in the standardisation of the Shona language. Then this has made the extension of the use of Shona possible in areas previously the reserve of English only. This in turn has empowered the Shona-speaking communities who can now use the language they understand

and identify with. What it means is that the majority of the people can now participate in developmental issues without being excluded, because of language.

Bibliography

- Barnes, E.W.** 1932. *A Vocabulary of the Dialects of Mashonaland in the New Orthography*. London: The Sheldon Press.
- Biehler, E.** 1950. *A Shona Dictionary with An Outline of Shona Grammar*. Revised edition. Chishawasha/Salisbury: The Jesuit Fathers / Cape Town: Longmans, Green.
- Chakamba, R. et al.** 1987. *Jekesa Pfungwa* 3. Harare: Longman.
- Chigidi, W.L.** 1988. *Focus Study Aids: A Level Shona*. Harare: College Press.
- Chimhundu, H.** 1997. Language Standardisation without Policy or Planning: Zimbabwe as a Case Study. Røyneland, V. (Ed.). 1997. *Language Contact and Language Conflict*: 129-150. Volda: Volda College.
- Chimhundu, H.** 2005a. *Lexicography and Language Raising: Dictionaries in Zimbabwean Languages*. Public lecture presented as part of the Golden Jubilee Celebrations of the University of Zimbabwe, Harare, 13 April 2005.
- Chimhundu, H.** 2005b. *Nhaka yaDoke*. Harare/Oslo: ALLEX Project/Department of Linguistics and Nordic Studies.
- Chimhundu, H.** 2005c. Doke and the Development of Standard Shona. Introduction to a photographic reprint of Doke, C.M. 1931. Harare/Oslo: ALLEX Project and Unipub AS.
- Chimhundu, H.** 2006. *Languages and Cultures of Marginalized Groups: Challenges of Inclusion in Higher Education*. Paper presented at the International Conference on the Transformation of Masvingo State University to Great Zimbabwe National University, Great Zimbabwe Hotel, 17-19 January 2006.
- Chimhundu, H. (Ed.)**. 1996. *Duramazwi reChiShona*. Harare: College Press.
- Chimhundu, H. (Ed.)**. 2001. *Duramazwi Guru reChiShona*. Harare: College Press.
- Dale, D.** 1975. *A Basic English-Shona Dictionary*. Gwelo: Mambo Press.
- Dale, D.** 1981. *Duramazwi: A Basic Shona-English Dictionary*. Gwelo: Mambo Press.
- Doke, C.M.** 1931. *Report on the Unification of Shona Dialects Carried Out under the Auspices of the Government of Southern Rhodesia and the Carnegie Corporation*. Hertford: Stephen Austin.
- Dube, S.** 1995. *A Study into the Reasons for the Diglossic Relationship between Shona and English in Post-independent Zimbabwe*. Unpublished B.A. Honours Thesis. Pretoria: University of South Africa.
- Fasold, R.** 1984. *The Sociolinguistics of Language*. Oxford: Basil Blackwell.
- Ferguson, C.** 1959. Diglossia. Giglioli, P. (Ed.). 1972. *Language and Social Context: Selected Readings*: 232-251. Harmondsworth: Penguin Education.
- Fishman, J.A.** 1972. *Language in Sociocultural Change*. Stanford: Stanford University Press.
- Fortune, G.** 1955. *An Analytical Grammar of Shona*. London: Longmans, Green.
- Fortune, G.** 1957. *Elements of Shona*. London: Longmans, Green.
- Fortune, G.** 1967². *Elements of Shona*. Salisbury: Longmans.
- Fortune, G.** 1972. *A Guide to Shona Spelling*. Salisbury: Longmans.
- Fortune, G.** 1984. *Shona Grammatical Constructions Volume 1*. Harare: Mercury Press.

- Fortune, G.** 1985. *Shona Grammatical Constructions Volume 2*. Harare: Mercury Press.
- Government of Zimbabwe.** 1987. *The Education Act, Chapter 55*. Harare: Government Printers.
- Hadebe, S. (Ed.)**. 2001. *Isichazamazwi SesinDebele*. Harare: College Press.
- Hannan, M.** 1959. *Standard Shona Dictionary*. Salisbury: The College Press.
- Hannan, M.** 1974. *Standard Shona Dictionary*. Second Edition. Salisbury: Rhodesia Literature Bureau.
- Hudson, R.A.** 1996. *Sociolinguistics*. Second edition. Cambridge: Cambridge University Press.
- Madzimbamuto, F.** 2006. Muzongozo neHungunyo (The Vertebral Canal and its Contents). Draft translation from Ellis, Harold, Stanley Feldman and William Harrop-Griffits. 2006. *Anatomy for Anaesthetics*. Eighth Edition. London: Blackwell.
- Magwa, W.** 1999. *Manyorerwo eChiShona*. Gweru: Mambo Press.
- Mashiri, P.** 2002. Shona–English Code-Mixing in the Speech of Students at the University of Zimbabwe. *Southern African Linguistics and Applied Language Studies* 20 (4): 245-262.
- Mashiri, P. and C. Warinda.** 1999. *Dudziramutauro reChiShona "A" Level*. Gweru: Mambo Press.
- Mheta, G. et al. (Eds.)**. 2005. *Duramazwi reMimhanzi*. Gweru: Mambo Press.
- Mpofu, N. et al. (Eds.)**. 2004. *Duramazwi reUrapi neUtano*. Gweru: Mambo Press.
- Mutangadura, J.** 2001. *The Language of Advertising: A Linguistic Approach to the Analysis of Advertisements*. Unpublished M.A. Dissertation. Harare: University of Zimbabwe.
- Nkomo, D. and N. Moyo (Eds.)**. 2006. *Isichazamazwi SezoMculo*. Gweru: Mambo Press.
- Nyota, S.** 1999. *Dudziramutauro reChiShona "O" Level*. Harare: Academic Books.
- Pongweni, A.J.C.** 1983. *The Language of Advertising and its Social Cultural Implications for the Consumer*. Issues in Development Seminar Series. Harare: University of Zimbabwe.
- Roy-Campbell, Z.M. and W.B. Gwete.** 1997. *Language Policy and Planning (ECS 207)*. Harare: University of Zimbabwe, Centre for Distance Education.
- Tsodzo, T.K.** 1992. *Rurimi Rwaamai Book 3*. Harare: College Press.
- Veit-Wild, F.** 2006. *Zimboicious — New Strategies for Verbal Exchange and Expressions in a Bilingual Cultural Setting: The Case of Zimbabwe*. Unpublished paper presented at a public lecture at the University of Zimbabwe, 26 August.
- Wild, H.** 1953. *A Southern Rhodesia Botanical Dictionary of Native and English Plant Names*. Salisbury: Government Printers.

Furthering the Aim of Multilingualism through Integrated Terminology Development*

Pumlani M. Sibula, *Unit for isiXhosa, Stellenbosch University Language Centre, Stellenbosch, Republic of South Africa (pmsibula@sun.ac.za)*

Abstract: This article aims at giving a brief account of how the activities of the Unit for isiXhosa, one of the five units of the Stellenbosch University Language Centre, contributes to the development of integrated technical terminology lists for all departments of the University. After the reason for this terminology development has been explained, a description is given of the six trilingual terminology lists that have been compiled so far. On a discussion of the problem of finding suitable isiXhosa terminological equivalents follows an exposition of the methodology used in compiling these lists. Finally it is shown how the collaboration with various bodies benefits the Unit for isiXhosa. The article concludes by stating how these lists have been received and are made known.

Keywords: MULTILINGUALISM, TERMINOLOGY DEVELOPMENT, EQUIVALENTS, SUBJECT AND LANGUAGE PRACTITIONERS, ACADEMIC LANGUAGE, TRANSLATION, TERM-FORMATION, TECHNICAL TERMINOLOGY, WORD LISTS

Opsomming: Bevordering van die doelwit van meertaligheid deur geïntegreerde terminologieontwikkeling. Hierdie artikel beoog om 'n kort verslag te gee van hoe die bedrywighede van die Eenheid vir isiXhosa, een van die vyf eenhede van die Taalsentrum van die Universiteit van Stellenbosch, bydra tot die ontwikkeling van geïntegreerde tegniese terminologieslyste vir alle akademiese departemente van die Universiteit. Nadat die rede vir hierdie terminologieontwikkeling verduidelik is, word 'n beskrywing gegee van die ses drietalige terminologieslyste wat reeds sover saamgestel is. Op 'n bespreking van die probleem om gepaste isiXhosa terminologiese ekwivalente te vind, volg 'n uiteensetting van die metodologie wat gevolg word by die samestelling van hierdie lyste. Laastens word aangetoon hoe die samewerking met verskillende liggame die Eenheid vir isiXhosa bevoordeel. Die artikel eindig deur te vermeld hoe hierdie lyse ontvang is en bekend gestel word.

Sleutelwoorde: VEELTALIGHEID, TERMINOLOGIEONTWIKKELING, EKWIVALENTE, VAK- EN TAALPRAKTISYNS, AKADEMIESE TAAL, VERTALING, TERMVORMING, TEGNIESE TERMINOLOGIE, WOORDELYSTE

* This article is based on 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

The University of Stellenbosch, being an Afrikaans institution in its tuition, is faced with the challenge of catering for other students, especially isiXhosa-speaking students, who have no previous knowledge of Afrikaans. In April 2003 the Stellenbosch University Language Centre established a Unit for isiXhosa, alongside the other five existing units, two of which are aimed at Afrikaans and English. The other three units are the following:

- The Unit for Document Design focuses on research on document design, text structure, text quality, professional communication, scientific writing skills and related topics.
- The Writing Laboratory takes a client-centred approach to create a space where all students and staff of the University can participate in academic discourse about writing.
- The Language Service offers translation, editing, copywriting, transcription and whisper interpreting services to clients on and off the University campus.

The two units dealing with Afrikaans and English have the following objectives:

- The Unit for Afrikaans chiefly aims at empowering students of the University to enable them to utilize generic language and thinking skills in a strategic and responsible manner within different contexts.
- The Unit for English aims at assisting students to acquire and polish reading, writing and speaking skills necessary for successful academic performance.

To these two units have been added another for isiXhosa:

- The Unit for isiXhosa functions in close co-operation with the Department of African Languages. It has the special task of contributing to the promotion of isiXhosa as an academic language, creating opportunities to further the use of isiXhosa in general. In this way the University of Stellenbosch has responded to the Constitution of South Africa, which provides for multilingualism and the development of the South African linguistic heritage. By setting the Unit for isiXhosa this task, the University wishes to contribute to the development of the indigenous languages of South Africa, IsiXhosa especially being the most important indigenous language in the Western Cape region.

The services rendered by the Unit for isiXhosa

The Unit for isiXhosa offers a twofold service to the realization of isiXhosa.

Firstly, it is responsible for the development of language skills in isiXhosa. In collaboration with the Department of African Languages, short acquisition courses are presented to both the University staff and students. These courses provide them with multilingual skills in communicating with isiXhosa speakers in their own language. Visiting overseas students gain access to these courses through the International Office, which promotes and supports the international mobility of academic staff and students by developing and implementing the University's international relations. These communication skills are especially beneficiary to non-South African students as their own institutions sometimes require them to learn something about the languages and cultures of the particular countries when they are doing projects among the communities there.

Secondly, the Unit for isiXhosa is responsible for the development of integrated technical terminology lists for the academic departments of the University of Stellenbosch. These lists are compiled with the main aim of providing academic support to students with isiXhosa as mother tongue so that some of their academic material can be learnt in their own language. Through this project, it is hoped that students will gain a better understanding in their study activities. This concurs with the remarks of Dlodlo (1999: 328): "The use of mother languages of instruction ... should facilitate the extension of educational opportunities in science and technology at colleges and universities, particularly to those students from rural communities whose command of the English language is limited." The fact that departments of African languages at South African universities are closing down because student numbers have fallen drastically, has been highlighted from time to time. This dwindling of numbers in other Southern African countries has also been noticed by Dlodlo (2005: 8), who emphasizes the importance of the development of African languages by saying that "our indigenous languages are dying — our languages must be used extensively to empower people".

IsiXhosa, belonging to the Nguni family of languages, is one of the official languages of South Africa and next to English and Afrikaans the most used language in the Western Cape. Apart from the Western Cape, it is spoken in five other provinces, namely: Eastern Cape, Northern Cape, Free State, North-West and Gauteng.

Together with other South African indigenous languages, isiXhosa is hampered by several factors which diminish its status. As Alberts (2000) indicates, the government, through its policies prior to 1994, marginalized the indigenous languages of South Africa, preventing them from gaining the same powerful status as English and Afrikaans. This state of affairs made its speakers lose confidence in isiXhosa as a language of wider communication, allowing themselves to be absorbed by other dominant languages. It is important to restore the people's confidence in their own language. This idea is supported by Hoffmann and Ytsma (2004: 32) who observe that language is the main instrument for upholding or promoting a group's ethnic identities.

The reason for terminology development

The isiXhosa terminology project aims at providing, especially for students, basic technical vocabularies in different subject fields through the translation of terms. In this way, they will acquire the necessary vocabulary needed to communicate scientifically in these particular subject fields. The compilation of these terminological lists will enable isiXhosa-speaking students to learn the basic concepts of these subject fields in their own language.

Sharing knowledge is one of the purposes of this project. The idea is to improve communication by enabling the target group to have this developed terminology at their disposal to designate the same concepts with the same terms. This will help to clear up the confusion and to avoid the misunderstanding caused by the use of different terms for the same concept.

Apart from the fact that, according to Hoffmann and Ytsma (2004: 62), language is a symbol of identity as well as a means of promoting ethnic identities, it is a resource of multilingualism which will give its speakers access to a greater body of knowledge and larger social circles. It is a kind of knowledge that will enable them to have a better understanding of other members of society through a language culture.

By using a sociolinguistic approach, this project has been developed to create new terms in isiXhosa, a development which reflects the influx of new ideas into the language to such a degree that, as community members, they have acquired a new way of looking at reality and a new self-awareness. It will forge a link between the community and the technical world as Cluver (1989: 199) suggests. "The aim of terminology development," says Alberts (2000: 238), "is to support ... the formulation, development, implementation of a national policy and strategies concerning technical languages with a view to the promotion of scientific and technical communication."

This will hopefully lead to determined efforts by the target groups of the different terminology lists to revitalize and revive their language, so that, in spite of its utilitarian function, it will also become a source of pride. In the environment of the University of Stellenbosch, where isiXhosa comes only third, after Afrikaans and English, these lists will regulate usage, mitigate misunderstanding, and improve communication.

It is the primary objective of this terminological project to assist isiXhosa students in understanding Afrikaans and English textbooks in their own language, which may result significantly in increasing their academic performance, because cognition and conceptualization will have taken place through their home language. This terminological project is accomplished in an integrated way, being performed in close collaboration with the different academic departments and faculties to ensure that the terminology in common use is provided. The Unit has already developed seven trilingual terminology lists, which can broadly be divided into two groups: (a) Generic terminology lists which comprise an Academic list, an Administrative list, and a list of Faculties, Departments and Academic Divisions together with Administration and Ser-

vices, and (b) Subject-specific terminology lists which include lists of Sociology, Social Work, Psychology and Law terms.

All these lists have been published in booklet format, with the exception of the list of Faculties, Departments, and Academic Divisions together with Administration and Services which is available electronically on the website of the Language Centre: <http://www.sun.ac.za/taalsentrum>. The reason for its electronic format is that it can easily be changed and added to when necessary. This is also the only list which does not contain short definitions in English, Afrikaans and isiXhosa of all the terms listed.

(a) Generic terminology lists

Academic list

Isigama sobuNgcali beMfundo/Generic Academic Terminology/Generiese Akademiese Terminologie comprises 147 entries of terms commonly used in the University environment. The terms were collected from various sources of an academic nature. This is the only list having a reference register at the back to guide the user in finding the English and isiXhosa equivalents for the Afrikaans terms. It was published in 2004 with C.B. Basson as co-compiler.

Administrative list

Isigama sobuNgcali boLawulo/Generic Administrative Terminology/Generiese Administratiewe Terminologie contains 206 terms which a student may encounter in the administrative environment of the University of Stellenbosch. The list is meant for isiXhosa students who do not understand Afrikaans and who have English as a second language. The main purpose for compiling this list was to help these students to find their way through the process of administration at the University. In the initial conceptualization of this list, the publication *WE-TUS: Wegwyser vir Engelse Taalgebruik aan die US*, compiled by Walter Winckler and Mattie van der Merwe, was used as a guideline. The list, co-authored with C.A. Basson, was published in 2004.

Before publication, the Academic and Administrative lists were submitted electronically to various subject and language specialists for evaluation and editing. Later, a workshop was organized for coordinating and considering the comments of these practitioners on these lists before they were finally adopted for publication.

Faculties list

This list which comprises 183 entries, is divided into two parts: first, *IiFakhalthi, amaSebe namaCadelo eMfundo/Faculties, Departments and Academic Divisions/Fakulteite, Departemente en Akademiese Afdelings* (with 147 entries), and second,

ULawulo neeNkonzo/Administration and Services/Administrasie en Dienste (with 36 entries).

Except for the terms taken from the calendar (prospectus) of the University of Stellenbosch, it lists the names of the faculties, departments, centres and other academic divisions collected from the University website where they are presented in both Afrikaans and English. To these the isiXhosa terms were added later. After compilation this list was made available electronically to all the faculties and departments, giving them the opportunity to comment on the isiXhosa equivalents provided. Most departments found these acceptable, with the exception of a few that preferred to use the transliterated equivalents. For example, the Faculty of Arts preferred to use *ifakhalthi yeAthsi* instead of *ifakhalthi yobuGcisa*, which was found not comprehensive enough to include all the departments falling under it. The Department of Logistics also preferred to use the transliterated term *iLojistiki* rather than any of the other names suggested.

(b) Subject-specific terminology lists

Sociology list

Isigama seSosiyoloji/Sociology Terminology/Sosiologie-Terminologie consists of 344 entries. The English entries together with their definitions were drawn from the third edition, published in 2004, of Margaret L. Andersen and Howard H. Taylor's book *Understanding a Diverse Society* with the permission of Wadsworth Publishing Company. The isiXhosa and Afrikaans equivalents were added later. The terminology presented is aimed at first-year students in Sociology. It was published in 2005 with Erina Basson as co-author.

Social Work list

Isigama sobuNtlalontle/Social Work Terminology/Maatskaplikewerk-Terminologie contains 325 entries. The isiXhosa version was added to the existing English and Afrikaans taken from the *New Dictionary of Social Work/Nuwe woordeboek vir maatskaplike werk* compiled by the Terminology Committee for Social Work. The terminology is meant for use by first-year students taking Social Work as subject.

Psychology list

Isigama esingengabula-zigcawu kwiSayikholoji/An Introduction to Psychology Terminology/'n Inleiding tot Sielkunde-Terminologie consists of 514 entries. The list is divided into nine chapters: psychology as a science and a profession, physiological psychology, sensation and perception, state of consciousness, learning, memory, intelligence, thinking and language, and motivation and emotion. The terms were initially drawn from the book *Psychology: An Introduction for Stu-*

dents in Southern Africa by D.A. Louw and D.J.A. Edwards, the second edition of which was published in 1997 by Heinemann. Both isiXhosa and Afrikaans equivalents were added later. The list is meant for first-year students in Psychology. The original list in English, compiled by Dr Judora Spangenberg, a former lecturer at the Department of Psychology, University of Stellenbosch, was reworked to its present form. Published in 2006, it was co-authored with Deirdré Roos.

Law list

Isigama sezoMthetho/Law Terminology/Regsterminologie comprises 283 entries divided into four chapters: family law, customary law, criminal law and an introduction to law. The family law terms were extracted by Prof. S. Human from the book *Introduction to South African Family Law* which she co-authored with J.A. Robinson and A. Boshoff and which was published in 2002. Ms M. Nel compiled the criminal law terms from the third edition of C.R. Snyman's book *Criminal Law* which was published in 1995 by Butterworth, while a senior student Phozisa Nana was responsible for the definitions. Phozisa Nana also provided the terms and definitions in English for the chapter on customary law. The terms for the chapter "Introduction to law" were extracted by Prof. Lourens M. du Plessis from his book *Introduction to Law* published by Juta. The terminology is aimed at first-year Law students. The Latin terms are presented untranslated, while the English terms and definitions are translated into isiXhosa and Afrikaans. The publication year is 2007 and the co-author Deirdré Roos.

The provision of equivalents

The isiXhosa equivalents provided in these lists are of a mixed nature. Indigenous equivalent terms were used in so far as they were available. Though there are some transliterated term equivalents, most of the equivalents consist of a brief translation of the source language terms so that there is a one-to-one meaning relation between the source language and target language equivalents. "Borrowing directly from European languages, by phonetic transcription," says Dlodlo (2005: 325), "should be avoided since such borrowed words do not convey any meaning initially and such a method produces clumsy-sounding Nguni science words." The syntax, word structure, vocabulary and sound system of English are so different from those of African languages that a useful scientific vocabulary cannot be created by transliteration. This agrees with Mphahlele (2001) who argues for the use of translated equivalents rather than transliterated ones. He supports his argument by saying that, because of the one-to-one meaning relation, the terminology user would be able to deduce the meaning of the source language terms.

However, sometimes it was impossible to find suitable translation equiva-

lents for concepts unknown in the isiXhosa vocabulary. As a result the terminologist was confronted with the challenge of transliterating the terms for these new concepts into isiXhosa. These adoptives were adapted and naturalized to be identified, written and articulated as isiXhosa words.

The compilation methodology

These terminology lists are compiled according to priorities determined by needs assessments. The requests from the departments are referred to the head of the Unit for isiXhosa who liaise with the head of that specific department to discuss the detailed specifications pertaining to the envisaged project.

When the director of the centre for funding has given the go-ahead for a project, the unit instructs the head to find a suitable person to provide a documented list of terms and their definitions in the source language, usually English. The list is handed to the Unit and the terminologists provide translation equivalents and definitions in isiXhosa and in Afrikaans. These terminological projects concur with the linguistic theory of Catford (1965) which states that a translation should have the same effect on a target language reader as the source language would have had on a source language reader.

The terminological list is then sent electronically to two subject-field specialists who are lecturers from the specific departments, one for isiXhosa and one for Afrikaans. They are responsible for quality control, giving their feedback in writing. After the necessary corrections have been made, these lists are sent to the language specialists in English, isiXhosa and Afrikaans for editing and proofreading the sections dealing with their specific languages, for checking the orthography, spelling, punctuation, etc. The edited lists are then entered into the Unit's database.

Finally these lists go to the Unit for Document Design, where they are converted into pdf files, the cover pages are prepared and then sent to the printers. When printed in booklet form, they are officially presented to the relevant departments. These departments then decide how these publications should be treated, for example whether they should be put on WebCT, as the Department of Sociology has done.

Collaboration with various bodies

The Unit for isiXhosa has effective partnerships and working relationships with various bodies. The nature of its work requires the Unit to have contractual ties for collaboration purposes, e.g. with departments and other institutions on the campus of the University of Stellenbosch. It has a networking relationship with the language specialists from the provincial and national bodies, who help through their professional service with translation, proofreading and editing.

One of the fruitful working relationships it enjoys is with the Pan South African Language Board (PanSALB). When the unit for isiXhosa hosted its first

terminology training in February 2006, the presenter Dr M. Alberts was from PanSALB.

The three-day workshop was held in close collaboration with the Department of African Languages at the University of Stellenbosch. Language practitioners of the indigenous South African languages were invited to speak on the principles, procedures and practice of terminology and terminography in the African languages.

The aim of the workshop was to equip language practitioners with all the basic skills needed to carry out a project through all the stages of the terminological process. It was a hands-on workshop that focused among others on the following topics: the cognitive, linguistic and communicative dimensions of terminology, word-forming principles, needs assessment, the consultation and planning of technical dictionaries, and the practice of term list compilation.

It was attended by participants from all over South Africa. It turned out to be a multilingual terminology training session, as the following languages were represented: Afrikaans, isiXhosa, isiZulu, isiNdebele, Sesotho, Setswana, and Sepedi. As a result of this diversity, the communication was in English.

The workshop proved to be a problem-solver for the participants who communicated the difficulties they encounter in their daily work as language practitioners. Dr Alberts's theoretical and practical sessions developed the participants' knowledge and expertise with regard to terminology development. They were all encouraged to discover and regard terminology development as a solution to language development.

The impact of the wordlists

The response to the wordlists compiled so far by the Unit for isiXhosa were positive throughout the University of Stellenbosch and even from other universities positive feedback was received. Especially the Departments at which the wordlists are aimed have done much to advertise these wordlists in specialized journals and on departmental web pages. They were also brought to the notice of dignitaries in specific government departments related to some of the subject fields which these lists cover.

The services and products of the Unit for isiXhosa are publicized and made known by means of brochures, bulletins and open days, and sold through effective marketing strategies.

References

- Alberts, Marietta.** 2000. Terminology Management at the National Language Service. *Lexikos* 10: 234-251.
- Catford, J.C.** 1965. *A Linguistic Theory: An Essay in Applied Linguistics*. London: Oxford University Press.

- Cluver, A.D. de V.** 1989. *A Manual of Terminography*. Pretoria: Human Sciences Research Council.
- Dlodlo, Temba S.** 1999. Science Nomenclature in Africa: Physics in Nguni. *Journal of Research in Science Teaching* 36(3): 321-331.
- Dlodlo, Temba S.** 2005. Teaching Science in Nguni Languages. *LEAPnews* 9: 8, December.
- Hoffmann, Charlotte and Jehannes Ytsma.** 2004. *Trilingualism in Family, School, and Community*. Clevedon: Multilingual Matters.
- Mphahlele, Motlokwe Clifford.** 2004. The Transliteration Principle: Is This the Best Procedure in African Language Lexicography and Terminology? *Lexikos* 14: 340-347.

On the Timelessness of Music Dictionaries

Henning Bergenholtz, *Department of Afrikaans and Dutch, University of Stellenbosch, Stellenbosch, Republic of South Africa; and Centre for Lexicography, Aarhus School of Business, Aarhus, Denmark (hb@asb.dk),*
and

Inger Bergenholtz, *Department of Music, University of Stellenbosch, Stellenbosch, Republic of South Africa; and School of Music, Kolding, Denmark (ibenholtz@mail.dk)*

Abstract: A music dictionary for the Internet fulfils the same functions as printed music dictionaries. An earlier music dictionary is as useful as a new one if its information is correct. But the fact that an Internet dictionary can at any time be corrected according to modern practices makes it, if not timeless, at least more up to date. Furthermore, the possibilities of illustrating with picture and sound open a wide field of usefulness. Nevertheless the lexicographer has to be aware of the different needs of different user types in different user situations. The dictionary being discussed, *Musikordbogen*, has been designed for text reception rather than translation or text production. After the inception of the dictionary has been described, the way the possibilities of the Internet has influenced the concept and the content of the articles and the outer texts is discussed.

Keywords: MUSIC DICTIONARY, DICTIONARY FUNCTIONS, SPECIALIZED LEXICOGRAPHY, INTERNET LEXICOGRAPHY, TRANSLATION, TEXT RECEPTION, TEXT PRODUCTION, MUSICAL KNOWLEDGE

Opsomming: Oor die tydloosheid van musiekwoordeboeke. 'n Musiekwoordeboek vir die Internet vervul dieselfde funksie as gedrukte musiekwoordeboeke. 'n Vroeëre musiekwoordeboek is net so nuttig soos 'n nuwe een indien sy inligting juis is. Maar die feit dat 'n Internetwoordeboek enige tyd gekorrigeer kan word volgens moderne praktyke, maak dit, indien nie tydloos nie, ten minste meer op datum. Verder open die moontlikhede van illustrasie met prent en klank 'n wye veld van bruikbaarheid. Nogtans moet die leksikograaf bewus wees van die verskillende behoeftes van verskillende gebruikertipes in verskillende gebruikersituasies. Die woordeboek wat bespreek word, *Musikordbogen*, is beplan vir teksontvangs eerder as vertaling of teksproduksie. Nadat die ontstaan van die woordeboek beskryf is, word die manier bespreek waarop die moontlikhede van die internet die ontwerp en inhoud van die artikels en die buitetekste beïnvloed het.

Sleutelwoorde: MUSIEKWOORDEBOEK, WOORDEBOEFUNKSIES, GESPECIALISEERDE LEKSIKOGRAFIE, INTERNETLEKSIKOGRAFIE, VERTALING, TEKSONTVANGS, TEKS-PRODUKSIE, MUSIEKKENNIS

1. The concept "timeless"

The title of this contribution about music lexicography is inspired by a headline in *Die Burger* of 26 March 2006: "WAT se waarde tydloos". A timeless dictionary would indeed help solving the problems of neither having enough dictionaries (cf. Bergenholtz and Gouws 2006), nor having the necessary money and manpower to update the completed dictionaries quickly enough. This headline was of course not written by a lexicographer who, more than anyone else, is aware of the changes and developments in the world and, dependent on these, in the languages of the world. Today a distinction between the music of Whites, Blacks and Coloureds as made in the foreword of *The South African Music Encyclopedia* (1979: i) would hardly be considered as essential for a music dictionary. And today a dictionary of molecular biology more than 15 years old would, for example, not be timeless because it would not describe the terms appropriately, due to the knowledge explosion in this field. In such a dictionary, there would be many serious lemma lacunae and even more incorrect lexicographical definitions. In other fields, one of these being music, similar explosive developments have not occurred. A music dictionary which is 100 years old will still be valid, if the information given in the entries is correct. It must come as a surprise to all ordinary dictionary users who regard their dictionary as infallible, that dictionary items are quite often wrong. The explanation for information errors in general language dictionaries concerning terms from specialized fields could be that they were not compiled by experts in the different fields. But this cannot be the only explanation, because the same and other wrong information is also found in specialized music dictionaries (often named *music lexicons* or *music encyclopaedias*). Nearly 50% of all terms in a Norwegian music dictionary from 1992 was faulty (see the review of Bergenholtz 1994). Almost all terms from the field of molecular biology in the big Norwegian national monolingual dictionary have incorrect lexicographical definitions, the explanations being not just incomplete, but outright erroneous (Bergenholtz 1996). Such errors are not only found in Norwegian dictionaries, they occur in dictionaries from most countries all over the world (cf. Bergenholtz 2007). Dictionaries from South Africa are no exception, as becomes clear from the following example from the *South African Music Dictionary* (2000: 156) where the abbreviations SH, A and E stand for Shona, Afrikaans and English respectively:

mbira [SH A E] (*Shona tokkel-idiofoon* • *Shona plucked idiophone*)

The mbira is an instrument employed in recreational activities and in ancestral communication. It is not only used by Shonas, but all over Southern and Eastern Africa, being called *mbira* or *ikembe* in different languages. This entry can therefore be justified by saying that the information is incomplete. This, however, is not the case with the following example from the *South African Music Dictionary* (2000: 156):

mechanical instrument, clockwork instrument [E]:

meganiese instrument [A]

meganiese instrument [A]: mechanical instrument,

clockwork instrument [E]

A clockwork instrument is a very special type of mechanical instrument. The two English terms are therefore not synonymous. A single Afrikaans term as equivalent for the double lemma is incorrect as translation for both terms; only the direct translational term is correct.

A similar misrepresentation can be found in the following example from the *South African Music Dictionary* (2000: 158)

metallofoon [A] (*Orff-instrument*): metallophone [E]

(*Orff instrument*)

Orff-instruments are often metallophones, but this is not the only possibility. Metallophones are used all over the world, from the gambang in the gamelan music of Indonesia, to the vibraphones of Western symphonic music. Such limited information can only result in a misleading lexicographical item.

In other words, a dictionary cannot be called timeless if it contains entries with incorrect information. Such errors which are common, can easily be prevented. It is, however, much more difficult to avoid the biggest problem for dictionary users: The entry not containing the needed items for the intended dictionary function, which is to help a translator rendering text correctly, a native speaker writing his mother tongue or a non-native speaker using a foreign language.

2. Dictionary functions

In the modern lexicographical theory of functions (cf. Tarp 1992, Bergenholtz and Tarp 2003), the following lexicographical functions are assumed, with which a dictionary must comply in order to meet the needs of the expected users:

communicative functions: text production, reception, translation

cognitive functions: systematic want for knowledge, sporadic want for knowledge, ...

By *communicative needs* is meant help in solving a problem, which may arise in relation to reading, producing or translating a text. By *cognitive needs* is meant help in achieving actual or systematic knowledge of something, which may be a linguistic or a non-linguistic phenomenon or context. The context of cognitive needs may be illustrated by the following sequence of events which tells nothing about why or in which context the user wants more knowledge:

dictionary user → dictionary → dictionary user

If the dictionary use is communication-related, a completely different sequence of events occurs. While reading, hearing, correcting, writing or translating a text, the user has come across a specific problem. Not knowing how to understand, correct, produce or translate a certain part of the text, he/she turns to a dictionary for help:

user → text part_z → user → dictionary → user → text part_z

This does not mean that there cannot be a cognitive gain through a communication-related use of a dictionary. This may be the case, but it is not the intention for the use of the dictionary.

The lexicographical function of a given dictionary is to provide assistance to a specific user group with specific characteristics in order to cover the complex of needs arising in a specific type of use situation. A dictionary can have one or more functions, i.e. it can be mono- or multifunctional. As any other utility product, a dictionary also has a genuine purpose, which consists of the totality of its functions and the subject field(s) it covers. Experience shows that the determinant in a dictionary function is the use situation. The difference between a dictionary designed for text production in the user's native language and one designed for translation into a foreign language is much greater than, for example, the difference between a native-language production dictionary designed for adult users and one designed for school children.

This fundamental issue seems to be recognized by the compilers of the *South African Music Dictionary* (2000: 5):

The function of the dictionary is mainly to **translate**, rather than to explain. The main aim is to promote the standardisation of Afrikaans and English terminology

This means that it is not the function of this dictionary to provide its users with special information about the subject field. If the user does not know what a *mbira* is, it cannot be learnt from this dictionary. In cases with correctly given equivalents, such a dictionary serves as a first aid, but according to all discussions in the field of specialized bilingual lexicography, this is only a small part of the needed information. If the user does not know the plural of *mbira*, he/she could guess that it should be *mbiras* in English, and in Afrikaans too, but he/she cannot be sure. The dictionary does not contain any grammar information, which is much needed for translation. And there are no collocational items, e.g. if the user needs to translate the English *pluck a mbira* into Afrikaans. Furthermore, when double terms are given, the user is not told which one is the correct standardized term. A dictionary is not always a dictionary. Sometimes it should be called a word list.

3. The concept of an Internet music dictionary

In one aspect it should be expected that an Internet dictionary would be timeless, especially if it is regularly updated, partly by assistance of user emails and

log-file analyses (cf. Bergenholtz and Johnsen 2005 and Prinsloo 2005). But this is not the case with printed dictionaries, which remain unchanged. It should be expected that printed dictionaries could be read many years after publication. This is not the case with electronic dictionaries. Almost all databased electronic dictionaries more than 10 years old can only partly, or at least not unproblematically, be read today if not regularly updated in the newest versions or transferred into other and more modern databases. This problem will not be dealt with here. Also, differences between electronic and printed dictionaries will not be discussed, except to say that both media try to fulfil the same dictionary functions.

The dictionary to be described is called *Musikordbogen*. It is an explanatory monolingual Danish music dictionary. The compilation was started during April 2005 in Denmark, was continued in Stellenbosch during the first part of 2006, and has, since its completion in August 2006, been accessible free of charge on the Internet at <http://www.musikordbogen.dk>. *Musikordbogen* (2006) is partly based on a printed dictionary, *Politikens musikordbog* (1995), which has a size similar to the *South African Music Dictionary* but contains fewer entries, about 3 000. However, it is no direct transformation from the printed to the electronic dictionary. All entries have been revised or rewritten and the systematic paragraphs on notation and music theory have been changed. There are many new entries. Until now, more than 500 lemmas have been added, as the length of an Internet dictionary has no financial implications. One of the main innovations consists of the dividing of the explanations to (a) a short definition/explanation, (b) a longer article with historical or other kinds of details and (c) an optional part with examples from music history. The printed version did contain references to other articles and the systematic paragraphs. In the Internet version, the access is made more easily as the user can just click on the references to switch from the article to the references and back. Apart from this, the Internet provides more varied possibilities as to illustrations, acoustic examples and www-links.

The intended functions of *Musikordbogen* can, on the one hand, be described as communication-orientated because the entries are meant to help laypersons and semi-experts with the reception of musical texts, be it music publications, radio programmes, musical notation etc. On the other hand, the systematic paragraphs contain overview articles on the principles of notation and music theory. Together with the more detailed parts of the explanations, they display the cognitive function of providing sporadic and systematic information about music for laypersons and semi-experts.

This dictionary is not meant for text production nor for translation. The translation of the cited articles and paragraphs would not have been possible by using the dictionary itself; a bilingual dictionary from and to English was employed.

The following description will give an idea of the structure of the dictionary. Obviously the actual layout of the entries seen on the Internet cannot be

shown. In practice, when having gained access to *Musikordbogen*, a field will appear where the sought-for term can be entered. To a certain degree, the system, like the system found on Google, will foresee misspellings, asking the user to select the right word. The first part of the entry can be shown separately. This may be considered as a help for the text reception function (translations are from the Danish (HB and IB)):

mbira

a small African idiophone consisting of metal strips fixed on a resonating box of bamboo or wood, to be plucked with the thumbs of both hands

If the user should want the cognitive function containing detailed information about the item, he/she can click on the complete article:

mbira

a small African idiophone consisting of metal tongues fixed on a resonating box of bamboo or wood, to be plucked with the thumbs of both hands

The mbira is also called *thumb piano*. The tongues may be made of wood, and the resonating box takes on many different shapes. The instrument is found in most of Southern and Eastern Africa. The mbira is the national instrument of Zimbabwe and is used by the Shonas in religious ceremonies. Sometimes it is found under the name of *ikembe* and it is often accompanied by a rattle named *hosho*, which is a calabash filled with seeds.

→ idiophone

= sansa

illustration



The article *idiophone* will appear after clicking on it. If an illustrative Internet address with a good picture of the instrument is found, hopefully with sound, it will appear in a separate field.

As mentioned above, a third part of the article will sometimes appear, mentioning examples from music history:

aleatoric

with no rigid pattern, only in a loose framework.

In the performance of aleatoric music, some of the rendition, or sometimes almost the whole of it, is left to the performing artist(s). The term was defined in 1955 and was immediately adopted by the composers attached to the so-called "Darmstädter Kurse". At first, the principle would only be effective as to the order of the movements, but later more and more of the elements were left to the musician(s). The composer John Cage (1912–92) was the first to use this concept under the name of *chance operation*, and after having developed it, he called it *indeterminacy*.

In 1985, John Cage composed the piano piece called *ASLSP (As SLow aS Possible)*, which he transferred to the organ in 1987. Apart from the tempo specification, the notation and the demand that at least one of the parts of the piece be repeated, everything else is left to the performer.

→ chance operation

→ indeterminacy

→ <http://www.john-cage.halberstadt.de/new/index.php?seite=galerie&l=d>

Sometimes the short explanation consists of only a translation of the term.

Synonyms are marked with an equality sign and clicking on them will show their articles:

ars vetus (lat.) the old art

This term belongs to the music theory of the 14th century. It is synonymous with *ars antiqua*, which is the more commonly used term.
= *ars antiqua*

Under some terms, the user will be guided to the systematic paragraph, which elaborates on the music theory concerned.

notation sign

the notation signs which are used to mark a change of pitch, i.e. the sharp, flat, double sharp and double flat and the natural.

The term can furthermore be used for the signs marking complications of the simple rhythms, such as the dot, the slur and the signs for triplets etc. Eventually the term is used for the signs which indicate that the notes be played at another pitch than notated, like *ottava bassa* etc.

→ accidentals

→ system of notation

→ <http://www.ccarh.org/courses/253/link/>

→ § 1

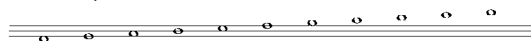
The systematic paragraphs consist of overview articles on the notation system and items from music theory, i.e. intervals, church modes, major and minor scales and their variants, the circle of fifths, and the structure of triads and other chords. At the end of the print version of the dictionary, illustrations of the normal arrangement of a Western symphony orchestra on stage and the range of the symphonic instruments and the human voice are given. These are substituted by references to Internet addresses and by pictures. Here again, the amount of information is no longer a problem, with no limitations on the size of the paper or the number of pages. This does not mean that there are no layout problems, but these should be left to experts on Internet design. Most of the systematic paragraphs comprise many illustrations in musical notation, which will appear as pictures in the running text. In the dictionary articles, they will be illustrations to be clicked on. If a term is given in bold letters, the user can click on the word, and the dictionary article will appear:

§ 1: Musical notation

As is outlined in the dictionary article **notation**, the art of writing music was developed in close connection to the music, which is to be reproduced. A direct result of this can be seen in the notation of early music, which was based on the normal dividing of the notes in three as a symbol of the Trinity. After the year 1600 when it can be said that church music had lost its dominating role, the notes were subdivided into two, whereas the dividing in three needs special signs. Especially after 1920, compositions with microtones and serial music showed the limitations of the so-called modern notation when being adapted to the diatonic scale system and to the harmonic structure of major and minor. In spite of great efforts to make the system clear, different periods and different persons will always add their own interpretations.

§ 1.1 System of notation

In modern musical notation, the notes are written on **staves** consisting of five lines. If the melody moves from one line to the next space, it has moved one step. This means that a scale of 11 notes and 10 steps can be written on one staff:



In a foreword to the systematic paragraphs, it is expressly mentioned that the paragraphs are not meant as a manual to music theory. If the user does not

know anything about the principles of Western music theory, he/she will not understand the text. But the paragraphs can help him/her to obtain an overview of the items, they can serve to elucidate forgotten facts, and the switching to and from the dictionary articles can assist in acquiring a better understanding.

4. Timeless dictionaries

Every dictionary is a document of its time; every printed dictionary is already dated when it has gone through the correction and printing stages. This can be avoided in an Internet dictionary because the lexicographer(s) can work on it until the moment it is presented to the public. There is another difference between printed dictionaries and Internet dictionaries: Internet dictionaries are quite often finished within a year or two; printed dictionaries are as a rule the result of many years' work. This does not necessarily mean that dictionaries finished in a short time are of a low quality. In fact, when many printed dictionaries, which have been worked on for many years are considered, some of them, on the contrary, contain a large number of serious errors. This may easily be explained by the fact that compiling a dictionary could be compared to the work of Schiller's bell founders: If the bell is made "in einem Gusse", i.e. in one casting, the result will be more satisfactory.

Literature

- Bergenholtz, H.** 1994. Hvorfor ikke skrive af? Anmeldelse af: Leiv Flisnes: *Musikordbogen*. [Oslo]: Tano 1992. *LexicoNordica* 1: 249-251.
- Bergenholtz, H.** 1996. Norsk international ordbog [review of the *Norsk ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. Band III. Flusker-Gigla. Edited by Reidar Bø, Arnbjørn Hageberg, Laurits Killingbertrø, Sigurd Norlie and Gunnar Pedersen. Oslo: Det Norske Samlaget 1994.]. *Norsk Lingvistisk Tidsskrift* 14: 75-84.
- Bergenholtz, H.** 2007. Falsche und richtige lexikographische Definitionen. Gottlieb, Henrik, Jens Erik Mogensen and Arne Zettersten (Eds.). *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 125-132. Amsterdam/Philadelphia: Benjamins.
- Bergenholtz, H. and R.H. Gouws.** 2006. How to Do Policy with Dictionaries. *Lexikos* 16: 13-45.
- Bergenholtz, H. and M. Johnsen.** 2005. Log Files as a Tool for Improving Internet Dictionaries. *Hermes* 34: 117-141.
- Bergenholtz, H. and S. Tarp.** 2003. Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. *Hermes, Journal of Linguistics* 31: 171-196.
- Brand, G.** 2006. WAT se waarde tydloos. Staatsubsidie egter weer ingekort. *Die Burger*, 25 March: 18.
- Musikordbogen* = Inger Bergenholtz in Cooperation with Richard Almind and Henning Bergenholtz. 2006. *Musikordbogen*. <http://www.musikordbogen.dk>.

Politikens musikordbog = Inger Bergenholtz. 1995. *Politikens musikordbog*. Copenhagen: Politikens Forlag.

Prinsloo, D.J. 2005. Electronic Dictionaries Viewed from South Africa. *Hermes* 35: 11-35.

South African Music Dictionary = Reino Ottermann and Maria Smit (Chief Eds.), assisted by Izak Grove, Winfried Lüdemann, Heinrich van der Mescht and Caroline van Niekerk. 2000. *Suid-Afrikaanse Musiekwoordeboek/South African Music Dictionary*. Compiled by a Revision Commission of the Suid-Afrikaanse Akademie vir Wetenskap en Kuns in Collaboration with the National Terminology Services. Second revised and enlarged edition. Cape Town: Pharos.

South African Music Encyclopedia = Jacques P. Malan (Gen. Ed.). 1979. *South African Music Encyclopedia. Volume I A–D*. Cape Town: Oxford University Press.

Tarp, S. 1992. *Prolegomena til teknisk ordbog*. Ph.D. dissertation. Aarhus: Department of Spanish. Aarhus School of Business. <http://www.lng.hha.dk/dml/spa/phd.pdf> [April 2006].

Implementing a Dictionary Culture in South Africa: An Attempt at a General Framework for the African Languages*

Juliane Klein, *Institut für Afrikanistik, University of Leipzig, Leipzig,
Germany (klein.juliane@freenet.de)*

Abstract: The development of a dictionary culture consists of three main phases: (a) the information phase, (b) the learning phase and (c) the consolidation phase. The first step towards a new dictionary culture is to inform mother-tongue speakers of available dictionaries and their benefits for individual speakers. This information should mainly be distributed through dictionary awareness campaigns conducted in the relevant African languages and by influential mother-tongue speakers. The main difficulty of the learning phase consists in the teaching of different groups, for example lexicographers, teachers and students, as well as adults who want to consult a dictionary effectively. Although a basic knowledge of lexicography is useful for teachers, a natural association with dictionaries as well as an intimate knowledge of dictionaries is more important. An intimate knowledge of dictionaries includes, among others: the different dictionary types, the contents of dictionaries and the know-how to find the sought-for word. The last phase is the consolidation phase in which the association with dictionaries will change from an uncommon to a normal activity.

Keywords: DICTIONARY CULTURE, INFORMATION, NATIONAL LEXICOGRAPHIC UNITS, DICTIONARIES, AFRICAN LANGUAGES, DICTIONARY AWARENESS CAMPAIGNS, EMPOWERMENT, TEACHING PHASE, USER, MEDIA, CONSOLIDATION PHASE, INCLUSIVE MODEL

Zusammenfassung: Implementierung einer Wörterbuchkultur in Südafrika: Ein Versuch eines allgemeinen Modells für die afrikanischen Sprachen. Die Entwicklung einer Wörterbuchkultur besteht aus drei Hauptphasen: (a) die Informationsphase, (b) die Lernphase und (c) die Festigungsphase. Der erste Schritt zu einer neuen Wörterbuchkultur ist es, Muttersprachler über erhältliche Wörterbücher und ihre Vorteile für den einzelnen Sprecher zu informieren. Diese Informationen sollen hauptsächlich durch Initiativen zum Gebrauch von Wörterbüchern, die in den entsprechenden afrikanischen Sprachen ausgeführt werden müssen, und einflussreiche Muttersprachler verbreitet werden. Die Hauptschwierigkeit der Lernphase besteht

* An earlier version of this article was presented as a paper at the Eleventh International Conference of the African Association for Lexicography organised by the Tshivenda National Lexicographic Unit, University of Venda for Science and Technology, Thohoyandou, Republic of South Africa, 5–7 July 2006.

darin, daß unterschiedlichen Gruppen zu unterrichten sind, z.B. Lexikographen, Lehrer, Schüler, sowie Erwachsene die ein Wörterbuch erfolgreich benutzen wollen. Obwohl ein grundlegendes Wissen über Lexikographie für Lehrer nützlich ist, sind der selbstverständliche Umgang mit Wörterbüchern sowie ein Grundwissen über Wörterbücher wichtiger. Zu dem Grundwissen über Wörterbücher zählen unter anderem: Die verschiedenen Wörterbuchttypen, Inhalte der Wörterbücher und das Wissen das gesuchte Wort zu finden. Die letzte Phase ist die Stabilisierungsphase in welcher der Umgang mit Wörterbüchern von einer seltenen zu einer alltäglichen Aktivität werden soll.

Stichwörter: WÖRTERBUCHKULTUR, INFORMATION, NATIONAL LEXICOGRAPHIC UNITS, WÖRTERBÜCHER, AFRIKANISCHE SPRACHEN, INITIATIVEN ZUM GEBRAUCH VON WÖRTERBÜCHERN, EMPOWERMENT, LERNPHASE, BENUTZER, MEDIEN, STABILISIERUNGSPHASE, INKLUSIVES MODELL

1. Introduction

In South Africa, mainly three groups are responsible for the compilation of dictionaries: (a) the National Lexicographic Units (NLUs), one for each official language, established by PanSALB, (b) the National Language Service (NLS) of the Department of Arts and Culture, and (c) commercial publishing houses and private individuals undertaking the production of general and technical dictionaries or word lists. General dictionaries, bilingual and monolingual, are compiled by the NLUs, whereas technical dictionaries and word lists are among the responsibilities of the NLS. Although all eleven official languages have at least some bilingual dictionaries, a developed dictionary culture has not yet been established for nine of the eleven languages.

There are two interdependent options to bring about a dictionary culture: (a) improve the skills of the dictionary users, and (b) improve the dictionaries themselves (Atkins and Varantola 1998: 83). Therefore, teaching the efficient use of dictionaries and compiling user-friendly dictionaries are the two most important challenges for lexicographers in South Africa. This article deals with the development of user skills, as "the majority of South Africans find themselves in a pre-dictionary culture environment" (Gouws and Prinsloo 2005: 42).

An officially multilingual country needs a dictionary culture, which includes at least all official languages of the country, i.e. eleven languages in the case of South Africa. Such a dictionary culture must be an inclusive model, based for example on the model of additive multilingualism or combined multiculturalism. This means that all user groups must be taken into account, because the South African dictionary users are not a homogeneous group. The different user groups include for example the general public, primary school pupils, secondary school pupils, teachers, and university students. Another feature which distinguishes the different groups is the existence of conjunctively written languages, i.e. the Nguni group (isiZulu, isiXhosa, Siswati and isiNdebele) and disjunctively written languages, i.e. the Sotho group (Sepedi, Sesotho, Setswana), Tshivenda and Xitsonga.

2. Implementing a dictionary culture in three steps

The development of a dictionary culture can roughly be divided into three phases: (a) informing people that dictionaries are available, (b) teaching people how to use dictionaries efficiently, and (c) accepting that the use of dictionaries is something normal, i.e. an everyday activity children learn in school as well as at home.

2.1 The information phase — Dictionary awareness campaigns

Informing members of a speech community about available dictionaries and their benefits is the first step towards establishing a new dictionary culture. Dictionary awareness campaigns form an essential part in this information phase and should consist of three stages: (a) Showing people which dictionaries are available, i.e. going to the speech communities, introducing them to the actual products and telling them where these can be found. (b) Illustrating the benefits of dictionaries for their users. (c) Explaining to people how to use a dictionary.

Showing people the actual dictionaries is important, as it can help to eliminate the fear of contact with such strange books if people can handle them and browse through them. Another advantage is that it can be explained to people how to use a dictionary and let them try it themselves afterwards. Explaining to people the way to consult a dictionary could be done in two steps: (a) Demonstrating to them how they can find the word they are looking for. (b) Explaining to them how to decode the information the dictionary entry contains, i.e. finding information about the meaning, usage, pronunciation, spelling, and grammatical behaviour of a word.

Illustrating the benefits of dictionaries aims to help people realize that dictionaries are a valuable means of empowering them in everyday life, enabling them to communicate with other people on a broad range of topics, i.e. providing them with the words they need to talk or write about topics in new domains. There are three areas of communication-oriented functions in which dictionaries help their users: (a) text production, (b) text reception, and (c) translation of texts (Gouws and Prinsloo 2005: 14). In other words, they help to extend the knowledge of the mother tongue and assist users in learning foreign languages.

Another indirect advantage dictionaries have for their users is their symbolic value, i.e., they show that the language is worth to be written down, and that it can be used in practically all domains of life. This has the positive side-effect that the status of the African languages is enhanced and that the L1 speakers of the African languages can be proud of their languages.

As the dictionary user is the most important factor deciding the success of a dictionary, a dictionary awareness campaign should be embarked on in accordance with the intended users. Questions to bear in mind concerning the users are: Who are the users? What do they know about the language? What do

they know about dictionaries? Do they know how to use dictionaries? It is essential that the language of such campaigns is a language the people understand. This implies that they should be conducted in the relevant *African language*, i.e. the African language of the dictionary, and not only in English, for only 25% of all South Africans understand English.

In the South African context, dictionary awareness campaigns could be designed and carried out by the dictionary compilers with the help and support of the relevant NLUs. Although this will be a difficult task for the NLUs, which do not have such large staffs and as much money as commercial publishers, it is currently the only possibility to promote their dictionaries. Another advantage is that the compilers are not only mother-tongue speakers, but also lexicographers. In order to minimize the cost of such campaigns, several NLUs could co-operate and share the costs of developing a marketing strategy which could be adapted to their particular languages, i.e. one concept for all, but adjusted according to the different languages.

The media could also play an important part in these campaigns in order to reach as many people as possible. A successful example of a broad media coverage of the launch of a new dictionary was that of the release of the SeDi-Pro (Sesotho sa Leboa Dictionary Project) online dictionary in South African newspapers during July 2003, which resulted in a significantly higher use of the dictionary than the official academic launch two weeks earlier (De Schryver and Joffe 2004: 190). This can only be successful if the media such as newspapers and radio stations are informed of available dictionaries by giving facts about the dictionaries through to book editors and programme managers and presenting them with review copies. This also means that the dictionary compilers themselves should reach out to the speech communities telling them of the completed dictionary, giving them some useful information about it (such as the intended user group, benefits for the user, and value for the speech community and the language), and showing them the product and, if possible, presenting them with a free copy.

Another psychological factor is the co-operation with influential South African institutions and people, who could support a dictionary awareness campaign for their L1 speakers. Ordinary people must be able to identify with such institutions and people, because they are respected and admired in a specific community. Except for churches and schools and their office-bearers, musicians, TV stars and sportsmen can be employed on a voluntary basis for introducing dictionaries to the general public.

In short, a dictionary awareness campaign ideally consists of three main parts: (a) telling people what is available and showing them the actual products, (b) illustrating to them the benefits of dictionary use, and (c) explaining to them how to use dictionaries efficiently.

2.2 Teaching dictionary use

The teaching phase faces three major obstacles: (a) Different groups are to be

taught, firstly, for example, dictionary compilers, lexicographic students, teachers, learners and students ranging from primary school to university, and, secondly, literate people who do not fit into any of these categories, but who also need training in consulting a dictionary. (b) At the moment, there are only a few skilled people who could do the teaching. (c) At first, all groups have to be taught simultaneously, which could lead to problems as the teachers especially must learn and at the same time teach what they are learning.

2.2.1 Instructing dictionary compilers and lexicographic students

Although there were some dictionaries compiled before the NLUs were established, the nine official African languages do not have a long dictionary tradition from which they could profit.

Examples of teaching lexicographic theory are workshops at certain lexicographic conferences and short courses at particular lexicographic units. At the Eleventh International Conference of AFRILEX a pre-conference workshop on "Principles and Practice of South African Lexicography" was presented. The Bureau of the WAT offers "a ten-day training course in general and computer lexicography and in the planning and management of a lexicographic project" (WAT 2005 online). In addition, the Department of Afrikaans and Dutch at the University of Stellenbosch have two postgraduate programmes on lexicography, an M.Phil. and a D.Litt. (SUN 2005 online). Such courses are valuable as they provide the first generation of properly trained lexicographers who are mother-tongue speakers of African languages.

2.2.2 Teaching teachers

Teachers, especially language teachers, but also subject teachers, should be able to use dictionaries confidently and knowledgeably and teach their pupils to consult dictionaries efficiently. Although a basic acquaintance with the theory of lexicography is useful for teachers, a confident use of dictionaries and a profound knowledge about dictionaries are more important. Because knowledge about dictionaries is a prerequisite for using dictionaries with confidence, it must therefore be taught to teachers. Such courses should be compulsory during teacher training and for all language teachers already in practice. They should include information on: (a) the benefits of dictionaries for the individual speaker, (b) the different dictionary types, (c) the contents of a dictionary, (d) the procedures to extract information from the dictionary, and (e) the methods to apply this information.

Being able to distinguish between the different dictionary types is important as they provide different kinds of information. Choosing the right type of dictionary is therefore vital for finding the desired information.

Knowledge about the contents of a dictionary concerns the structure of a dictionary, i.e. the difference between the central word list (what lay people call

"the dictionary") and the front and back matter, which provide information about the dictionary.

The next step is to extract the correct information from the dictionary. This means that teachers must know the alphabetical structure of the dictionary to find the required lemma and the construction of the dictionary articles to decode the information in them. As this can be difficult, especially for conjunctively written languages, the lemmatization principle of the dictionary must be clear, i.e. it must be certain what counts as a lemma and in which cases derivational forms are listed as separate lemmas. Breaking the words down into their components in order to find the correct lemmatized stem is usually easier for L1 speakers with mother-tongue intuition than for learners who are unsure to which noun or verb category a word belongs. Electronic dictionaries avoid this problem by including all forms and guiding the user directly to the correct lemma. Finding a solution to this problem for paper versions of the dictionaries will be an important step towards successful dictionary use. A first step in solving this problem is that dictionary compilers should state which lemmatization approach was used and which cases were considered as exceptions, explaining these clearly in the usage guide in the front matter of the dictionary. This leads to the next important aspect, which must be included in such a course: the use of the front and back matter of the dictionary. As these contain important information about the structure and contents of the dictionary and, ideally, a usage guide, they should not be ignored but considered as vital a part as the central word list.

Another important aspect to be treated in training teachers is illustrating ways to introduce dictionary use in the classroom. This is of special importance for language teachers who could use dictionaries in activities such as text reception, text production or translation.

2.2.3 Teaching pupils

The aim of teaching effective dictionary use, i.e. finding with confidence the correct information in the dictionary, is the same for all groups. Nevertheless, different approaches are needed for the different languages, age groups, as well as for L1 speakers and language learners. Teaching dictionary use should start as early as possible to ensure that pupils internalize it in their formative years. Dictionary use should not be taught as a single module, i.e. a module learnt for an examination and forgotten afterwards, because this will not lead to a regular consultation of dictionaries. Instead, the dictionary should be introduced as a useful tool helping students to communicate. It should be integrated on a regular basis in text decoding (reading) as well as text production (writing, speaking and word games), because practising the use of dictionaries is the only way of coming into the routine of using them. Starting to teach dictionary use in primary school implies that there must be school dictionaries available which suit the age of the users. This means they must be easy to use and must be written in a language the pupils understand in order to help them

to improve their communication skills (Gouws and Prinsloo 2005: 51). It is an important task for lexicographers to compile dictionaries for all age groups and especially for younger pupils, as they will be the ones who will be taught from primary school onwards how to use dictionaries. They could be members of the first generation of a new dictionary culture.

2.2.4 Teaching everybody else who needs to consult a dictionary

Teaching dictionary use is not only restricted to pupils, but must involve the whole speech community, i.e., everybody who may need to consult a dictionary. In speech communities where there is not an established dictionary culture yet, this will be very difficult as most of the speakers have already finished their formal education and cannot easily be reached. For these people, teaching dictionary use could mainly be done through dictionary awareness campaigns and the dictionaries themselves. This means that the front and back matter of dictionaries, and especially the user guides, are very important, because they will influence the success of the user. If the guide clearly explains how to find a word, i.e. if the lemmatization principle is clear and illustrative examples are given, users of the dictionary will, hopefully, find the meaning and use of the unknown word which they have come across.

2.3 The consolidation phase — The next generation

The last phase in establishing a dictionary culture will be the consolidation phase in which the use of a dictionary will become a normal, everyday activity for most members of the speech community. Or, as Jackson (2002: 21) puts it, "the dictionary (will) become part of the cultural fabric of (the) society". In this phase, the speech community takes over so that a dictionary culture will develop and grow through a system of cumulative distribution. Those pupils who will learn the consultation of a dictionary from an early age (ideally from primary level) onwards, will be the foundation on which this consolidation phase is built as they will have cultivated the use of a dictionary, teaching it to their children as a useful and informative procedure.

During this phase, there is a switch from a top-down to a bottom-up approach, i.e. the NLUs would not be able to control and direct the scene completely as the speech community will have more influence.

This phase includes not only a regular use of dictionaries in education, but also a feedback on the dictionaries by their users. This means that the speech community must be able to contact the lexicographers, giving them their opinions and experiences. Examples would be the feedback forms of online dictionaries (cf. SeDiPro) or any written or oral communication between the lexicographers and the users. A successful dictionary culture can only be established if the members of the speech community accept the dictionaries as something useful and informative in their own lives.

3. Conclusion

A successful dictionary culture needs an ongoing co-operation between lexicographers, who are compiling the dictionaries, on the one hand, and teachers, who are a vital factor in developing a dictionary culture, on the other. For South Africa, an inclusive dictionary culture is called for, i.e. a *multilingual* one comprising all official languages. This means that the theoretical concept is the same for all languages, but the actual models must be modified according to the needs of each speech community and each language. This aim could be reached through a strategy in which two approaches are applied simultaneously. The first is a broad short-term measure (the formation phase) that tries to reach as many people as possible and familiarize them with dictionaries through dictionary awareness campaigns. The second is a long-term measure (the teaching phase) that would not reach as many people as the dictionary awareness campaigns in the first years, but will produce a generation who has internalized dictionary use, seeing it as a normal, everyday activity.

Acknowledgement

I am grateful to Dr F. Rautert and Prof. E. Wolff for a very helpful and constructive critique of an earlier version of this article. However, they obviously carry no responsibility for any part of it.

Literature

Dictionaries

SeDiPro. <http://africanlanguages.com/sdp/>.

Other literature

Atkins, S. and K. Varantola. 1998. Monitoring Dictionary Use. Atkins, S. (Ed.). 1998. *Using Dictionaries*: 21-82. Tübingen: Max Niemeyer.

De Schryver, G.-M. and D. Joffe. 2004. *On How Electronic Dictionaries Are Really Used*. <http://www.tshwanedje.com>.

Gouws, R. and D. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.

Jackson, H. 2002. *Lexicography: An Introduction*. London: Routledge.

SUN (2005 online) <http://academic.sun.ac.za/afndl/programme/index.htm>.

WAT (2005 online) <http://www.sun.ac.za/wat/Engelse%20Webwerf/Training/WFB2.HTM>.

Kinship Terminology: Problems in Some English–Tshivenda Bilingual Dictionaries

Munzhedzi James Mafela, *Department of African Languages, University
of South Africa (UNISA), Pretoria, Republic of South Africa*
(mafelmj@unisa.ac.za)

Abstract: With the introduction of a new political dispensation in South Africa, bilingual dictionaries have become a necessary part of daily life. The main function of a bilingual dictionary is to provide one or more equivalents in the target language for the entry words in the source language. Not all bilingual dictionaries realize this objective satisfactorily, Tshivenda bilingual dictionaries included. Some bilingual dictionaries in Tshivenda provide one equivalent instead of all equivalents associated with the entry word of the source language, especially if Tshivenda is the target language. Not providing all equivalents renders a bilingual dictionary inadequate, because users do not always find what they look for. By focusing on kinship terms in some English–Tshivenda bilingual dictionaries, this article discusses examples of inadequate treatment. The article further advocates the inclusion of all possible equivalents under such entry words.

Keywords: BILINGUAL DICTIONARY, CROSS-RELATIONSHIP, DICTIONARY USER, EQUIVALENT, INDIGENOUS LANGUAGE, KINSHIP TERMINOLOGY, ORTHO-RELATIONSHIP, SOURCE LANGUAGE, LEXICAL UNIT, LEXICOGRAPHER, RELATIVES, TARGET LANGUAGE, TRANSLATION

Opsomming: Verwantskapsterminologie: Probleme in sommige Engels–Tshivenda tweetalige woordeboeke. Met die invoering van 'n nuwe politieke bedeling in Suid-Afrika het tweetalige woordeboeke 'n noodsaaklike deel van die daaglikse lewe geword. Die hoof funksie van 'n tweetalige woordeboek is om een of meer ekwivalente in die doeltaal vir die trefwoorde in die brontaal te verskaf. Nie alle tweetalige woordeboeke verwesenlik hierdie doel bevredigend nie, Tshivenda tweetalige woordeboeke ingesluit. Sommige tweetalige woordeboeke in Tshivenda verskaf een ekwivalent in plaas van al die ekwivalente wat in verband staan met die trefwoord van die brontaal, veral as Tshivenda die doeltaal is. Deur nie alle ekwivalente te verskaf nie maak 'n tweetalige woordeboek ontoereikend omdat gebruikers nie altyd kry wat hulle soek nie. Deur op verwantskapsterme in sommige Engels–Tshivenda tweetalige woordeboeke te fokus, bespreek hierdie artikel voorbeelde van onvoldoende behandeling. Verder bepleit die artikel die insluiting van alle moontlike ekwivalente onder sulke trefwoorde.

Sleutelwoorde: BRONTAAL, DOELTAAL, EKWIVALENT, INHEEMSE TAAL, KRUIS-VERWANTSKAP, LEKSIKALE EENHEID, LEKSIKOGRAAF, ORTOVERWANTSKAP, TWEETALIGE WOORDEBOEK, VERTALING, VERWANTE, VERWANTSKAPSTERMINOLOGIE, WOORDEBOEKGEBRUIKER

1. Introduction

A bilingual dictionary deals with the translation of entry words from one language into another language. Equivalents of the target language are provided for the entries of the source language. According to Zgusta (1971: 294), the aim of a bilingual dictionary is to co-ordinate the lexical units of one language with those units of another language which are equivalent in their lexical meaning. Whereas the basic function of a dictionary is to define words, the major task of a bilingual lexicographer is to find appropriate equivalents in the target language for units of the source language (Al-Kasimi 1983: 58). People daily need bilingual dictionaries for their translation problems and queries, especially those who learn other languages. Bilingual dictionaries are the most used in the new South Africa. Translation from languages such as English and Afrikaans into indigenous languages plays an important role, because all indigenous languages, alongside English and Afrikaans, are now recognized as official languages. People are entitled to receive information through their principal languages, hence the importance of translation. Bilingual dictionaries are therefore indispensable in the translation process.

Kinship terminology involves terms people in a society use to classify their relatives. It can therefore be defined as the system of names applied to categories of kin standing in relationship to one another (McHenry 1992a: 881). Stone (2000: 5) sees kinship as relationships between persons based on descent or marriage. However, kinship terminology varies from one language to another. In some instances, providing the actual meaning and correct equivalents of kinship terms is difficult. Languages differ in the way they express certain meanings. The target language may make more or fewer distinctions in meaning than the source language (Baker 1992: 22). In certain cases, the difference in the manner languages convey meaning is caused by cultural differences between the source language and the target language. McHenry (1992b: 282) stresses that "the finding of exact equivalents is more difficult than is commonly realized, because every language slices up the world in its own particular way". This is also expressed in kinship terminology. Stone (2000) says societies vary in the extent to which kinship connections form the basis of their social, economic and political structures. There may therefore not be a correspondence in terms. According to Al-Kasimi (1983: 63), two related items in two different languages may not cover the same semantic range.

In many instances, a kinship term in English has two or more equivalents in Tshivenda. Although Tshivenda has two or more equivalents for one kinship term in English, these are not reflected in some English–Tshivenda bilingual dictionaries. Only one equivalent is given, which may mislead a bilingual dictionary user, because all the meanings are not covered. As stated above, every culture has relations of kinship and has a system of kinship classification. The fundamental distinction running through Tshivenda kinship is that between *ortho-relationship* (a relationship between those of the same sex) and *cross-rela-*

tionship (a relationship between those of opposite sexes) (Van Warmelo 1904: 108). The purpose of this article is to highlight the lack of providing all the Tshivenda equivalents with regard to kinship terms in some English–Tshivenda bilingual dictionaries, and further to advocate that all the meanings of kinship terms be included as Tshivenda equivalents. The dictionaries from which illustrative material is drawn are the *English–Venda Vocabulary* by L.T. Marole and the *Improved Trilingual Dictionary: Venda–Afrikaans–English* by P.J. Wentzel and T.W. Muloiwa. The *Improved Trilingual Dictionary: Venda–Afrikaans–English* comprises three sections, namely, Venda–Afrikaans–English, Afrikaans–Venda, and English–Venda. In this discussion, the focus will be on the English–Venda section. Although these dictionaries are not as comprehensive as would be expected, they are better structured than the other English–Tshivenda bilingual dictionaries.

2. Equivalence and kinship terms in English–Tshivenda bilingual dictionaries

Prinsloo and De Schryver (2002: 162) define *equivalent* as a word or phrase in one language which corresponds in meaning to a word or phrase in another language. The issue of providing Tshivenda equivalents of the English kinship terms is one of the core problems in English–Tshivenda bilingual dictionaries. In English, the term *aunt* means the sister of one's mother or father, and the wife of one's uncle. However, in Tshivenda, *aunt* is used to denote a sister of one's father only, and she is called *makhadzi*. A sister of one's mother is called *mmane* (little mother), if younger than one's mother, and *mmemuhulu* (big mother), if older than one's mother. The adjectives *little* and *big* have been used to define the type of mother, because there are no English equivalents for *mmane* and *mmemuhulu* respectively. The wife of one's uncle is called *makhulu*. For the English kinship term *aunt* in this case, Tshivenda has three equivalents, i.e. *makhadzi*, *mmane* and *mmemuhulu*. In Marole (1954), only one equivalent is provided, i.e.

aunt — (n.) *makhadzi, khaladzi ya khotsi*

The equivalent provided in the entry is *makhadzi*, which is further defined as *khaladzi ya khotsi* (sister of one's father). The other meanings, *mmane* and *mmemuhulu*, have been omitted. This means that aunts from the side of the mother are not given. The failure of the lexicographer to provide the other two equivalents in his definition of the entry word *aunt*, may lead to dictionary users acquiring wrong meanings if they are looking for *mmane* and *mmemuhulu*. In the Tshivenda system, a person's mother and mother's sisters are called by one term and the father's sisters by another. One's mother and mother's sisters are called *mme* (mother), while the father's sisters are called *makhadzi*. The same applies to the kinship term *uncle*. In English, *uncle* means brother of one's father or mother, and an aunt's husband (Allen 1990). Scheffler and Lounsbury (1971:

7) say: "The English kinship term 'uncle' is polysemic; it may be used to designate the class of kinsmen defined as 'the brother of either of one's parents' or to designate one of two different but related classes of kinsmen, namely (1) the husbands of 'aunts' in which case we may speak more precisely of 'uncles by marriage' and (2) the brothers of grandparents, rather than of parents, in which case we may speak more precisely of 'great uncles.'" Polysemy refers to instances where the same word may have a set of different meanings. In Tshivenda, the use of *uncle* is restricted to the brother of one's mother and is referred to as *malume*. The brother of one's father is called *khotsimunene* (little father), if younger than one's father, and *khotsimuhulu* (big father), if older than one's father. A person distinguishes between the brothers of his/her father as *khotsimuhulu* or *khotsimunene*, depending on whether they are older or younger than one's father. In its kinship system, Tshivenda makes a distinction with regard to seniority. Brothers of a person's mother are called *malume* irrespective of relative age or status (Mönnig 1978: 241). As Hammel (1966: 7) says: "The relationship of seniority and juniority are, in this analysis, more important than the raw data would immediately indicate." The distinction by relative age is also applied to the sisters of one's mother, i.e. *mmane* and *mmemuhulu*. Marole (1954) provides one equivalent, *malume*, for the English kinship term *uncle*.

uncle — (n.) *malume*

In contrast to the equivalent *makhadzi*, there is no further definition. To make the meaning of *malume* clear to the dictionary user, the lexicographer should further have added *khaladzi ya mme* (brother of one's mother). Tshivenda indicates *father's brother* by one term, and *mother's brother* by a different one. The equivalents omitted in the definition are *khotsimunene* and *khotsimuhulu*. Uncles from the side of the father have been excluded. The exclusion of other equivalents may lead to a wrong choice of meaning of the English kinship term *uncle*. If the lexicographer had provided all the Tshivenda equivalents of the entry word *uncle*, he would have given dictionary users a wide scope of the meanings in the target language, from which to choose.

In English, *brother* designates "a male as he is related to other children of his parents" (Guralnik 1981). Tshivenda has three equivalents for the term *brother*, i.e. *mukomana* (elder brother), *murathu* (younger brother), and *khaladzi* (girl's male sibling). For *khaladzi*, age and gender are irrelevant. In Marole (1954), the equivalent of *brother* is given as follows:

brother — (n.) *murathu*

The equivalent only refers to one meaning of *brother*, namely *murathu* (little brother). Out of the three possible equivalents, only one has therefore been provided. The other two, *mukomana* (big brother) and *khaladzi* (girl's male sibling) have been omitted. As indicated above, leaving out other equivalents will lead dictionary users to make a wrong choice if the equivalent *murathu* is not required. The same applies to the English kinship term *sister*. Guralnik (1981)

defines *sister* as "a female as she is related to other children of her parents". For the English *sister*, Tshivenda has three equivalents, i.e. *mukomana* (elder sister), *murathu* (younger sister) and *khaladzi* (boy's female sibling). As in the case of the kinship term *brother*, the emphasis is on seniority. Larson (1984: 83) writes: "There are languages in which one cannot simply say *brother* because there may be two or more words to choose from." For both *brother* and *sister*, Tshivenda has three equivalents each. In Marole (1954), the Tshivenda equivalent of *sister* is provided as follows:

sister — (n.) *khaladzi wa tshisadzini*

The equivalent in Tshivenda is *khaladzi* with a qualifying phrase *wa tshisadzini* (who is female). The other two equivalents, i.e. *mukomana* and *murathu*, have been omitted. Definitions of the terms *brother* and *sister* in the target language are therefore not complete without the other equivalents.

The provision of the Tshivenda equivalents for the kinship terms *uncle*, *aunt*, *sister* and *brother* in Wentzel and Muloiwa (1982) has been handled fairly well. All meanings entailed in each kinship term have been provided. Dictionary users do not encounter problems in choosing the equivalents. The definitions of the English terms appear as follows in Wentzel and Muloiwa (1982):

uncle: (mother's brother) *malume (vho-)*; (father's elder brother) *khotsimuhulu (vho-)*; (father's younger brother) *khotsimunene (vho-)*

aunt: (father's sister) *makhadzi (vho-)*; (mother's elder sister) *mmemuhulu (vho-)*; (mother's younger sister) *mmane (vho-)*

sister: (sister of male person) *khaladzi*; (younger sister of female person) *murathu (vha-)*; (elder sister of female person) *mukomana (vha-)*

brother: (elder brother of male) *mukomana (vha-)*; (younger brother of male) *murathu (vha-)*; (brother of female person) *khaladzi (dzi-)*

The term *khaladzi* is used for a male or female sibling, irrespective of age, whereas the terms *murathu* and *mukomana* are used for a younger male or female sibling of the same sex and for an older male or female sibling of the same sex respectively. The prefixes *vho-*, *vha-* and *dzi-* in the definitions denote the plural forms of the Tshivenda kinship terms.

In English, children of one's father's or mother's brothers or sisters are one's cousins. This is not so in Tshivenda, because they are all considered one's brothers or one's sisters. According to Mönnig (1978: 237), the children of a father's brother are treated in much the same way as are a person's male siblings. In English, *cousin* means the son or daughter of one's uncle (including one's father's younger and elder brothers) or aunt (including one's mother's younger and elder sisters). In Tshivenda, the meaning of *cousin* is restricted to the children of one's father's sister and those of one's mother's brother. They are all referred to as *muzwala*. In his definition of the kinship term *cousin*, Marole

(1954) provides the equivalent as:

cousin — (n.) *muzwala*

Equivalents pertaining to children of the father's brother and those of the mother's sister are not provided, because in Tshivenda they are not considered as cousins, but as *mukomana*, *murathu* and *khaladzi*. A person learning Tshivenda will never know whether some cousins in Tshivenda are called *murathu* or *mukomana* or *khaladzi*, if the equivalents are not reflected. The lexicographer should provide the other meanings of *cousin*, because there is no other place where the dictionary user will find them, especially if he/she is learning Tshivenda. Wentzel and Muloiwa (1982) also provide the equivalent of the English kinship term *cousin* as *muzwala*, presenting it as follows:

cousin — (cross-cousin) *muzwala (vha-)*

As in Marole (1954), only cousins of cross-relationship are mentioned here. However, Marole does not indicate in his definition whether the term *cousin* involves cross-relationship or ortho-relationship. In both dictionaries, nothing is mentioned about cousins of ortho-relationship. If the dictionary user is looking for a cousin of ortho-relationship, he/she would be left in the lurch, because there is no way of finding it if not defined under the entry word *cousin*. In this case, a bilingual dictionary does not fulfil its purpose.

On the other hand, the kinship term *niece* means the daughter of one's brother or sister, and the daughter of one's brother-in-law or sister-in-law. In Tshivenda, *niece* denotes one's brother's daughter, if one is female, and one's sister's daughter, if one is male. If one is female, the daughter of one's brother-in-law is one's daughter, and if one is male, the daughter of one's sister-in-law is also one's daughter. The equivalent of *niece* in Marole (1954) is provided as:

niece — (n-) *muḍuhulu wa musidzana*

In this definition, the equivalent is given as *muḍuhulu* accompanied by a qualifying phrase *wa musidzana* which means "who is a girl". The lexicographer did this, because he knows that there is more than one equivalent for *niece* in Tshivenda. However, he did not provide the other equivalents, i.e. one's brother's daughter, if one is male, and one's sister's daughter, if one is female, because he is aware that in Tshivenda they are all called one's daughters. Furthermore, *muḍuhulu* does not take gender in consideration, both *nephew* and *niece* are called *muḍuhulu* in Tshivenda. For the lexicographer to help a dictionary user to discriminate the meanings, all senses must be reflected under the entry word *niece*. Wentzel and Muloiwa (1982) give the equivalents of *niece* as follows:

niece — *muḍuhulu (vha-); ṛwana (vha-)*

The lexicographers have provided two equivalents, i.e. *muḍuhulu* and *ṛwana*.

Each of the Tshivenda terms *muḏuhulu* and *nwana* has more than one meaning. The lexicographers should have discriminated the meanings of both terms to help dictionary users to make the correct choice of equivalents. For example, in Tshivenda *nwana* means one's own child, one's brother's child, if one is male, one's sister's child, if one is female, one's sister-in-law's child, if one is male, and one's brother-in-law's child, if one is female. In Tshivenda, the meanings of *muḏuhulu*, the equivalent of *niece*, are one's sister's child, if one is male, and one's brother's child, if one is female. It does not, however, denote one's own child. The equivalents provided should therefore read as follows:

- niece — *muḏuhulu* (vha-) granddaughter; one's sister's daughter — if one is male; one's brother's daughter — if one is female
- *nwana* (vha-) one's own child; one's brother's child — if one is male; one's sister's child — if one is female; one's sister-in-law's child — if one is male; one's brother-in-law's child — if one is female

The same holds for *nephew*, which is not included in Marole's bilingual dictionary. Wentzel and Muloiwa (1982) translate *nephew* as:

- nephew: *muḏuhulu* (vha-); *nwana* (vha-)

The Tshivenda equivalents of the English term are the same as those of *niece*. As with *niece*, the equivalents are *muḏuhulu* and *nwana*. The lexicographers did not discriminate the meanings encompassed by the equivalents *muḏuhulu* and *nwana*. As a result, dictionary users can encounter problems regarding obtaining the correct equivalent for which they are looking. For the purpose of including all the meanings of the equivalents, the definition of the entry word should look as follows:

- nephew — *muḏuhulu* (vha-) grandson, one's sister's son — if one is male; one's brother's son — if one is female
- *nwana* (vha-) one's own child; one's brother's child — if one is male; one's sister's child — if one is female; one's sister-in-law's child — if one is male; one's brother-in-law's child — if one is female; one's son-in-law

In contrast to English, the Tshivenda equivalents of *niece* and *nephew*, which are *muḏuhulu* and *nwana*, are not affected by gender. While in English the female is called *niece* and the male is called *nephew*, in Tshivenda both the female and the male are referred to as *muḏuhulu* and *nwana*. Unlike the other kinship terms discussed above, meaning discrimination plays an important role in the definitions of the Tshivenda equivalents *muḏuhulu* and *nwana*, because each equivalent encompasses more than one meaning. The absence of meaning discrimination will confuse dictionary users so that they might choose wrong equivalents.

3. Conclusion

The above exposition has shown that in the discussed English–Tshivenda bilingual dictionaries, the lexicographers did not achieve the objective of a bilingual dictionary, i.e. that of co-ordinating the lexical units of one language with those units of another language which are equivalent in their lexical meaning. The lexicographers failed to provide all possible Tshivenda kinship equivalents for the relevant English entry words. As a result, these bilingual dictionaries are not user-friendly. The discussion has also revealed that kinship terminology in English and Tshivenda differs, because kinship relationships are closely related to the speaker's experience. As Mbananga (2003: 115) writes: "The languages of different communities in South Africa are part of cultural diversity. Language has an intrinsic value closely connected with the cultural identity, knowledge and understanding." In showing the differences of kinship relation between English and African languages, Mandela (1995: 10) says: "In African culture, the sons and daughters of one's aunts and uncles are considered brothers and sisters, not cousins. We do not make the same distinctions among relations practised by Whites. We have no half-brothers or half-sisters. My mother's sister is my mother; my uncle's son is my brother; my brother's child is my son, my daughter." Bilingual dictionary lexicographers must always keep in mind that there will be no one-to-one correspondence between words across languages, because of cultural differences. Therefore, there is a need for lexicographers to acquaint themselves with the culture of the target language to provide all the possible equivalents.

References

Dictionaries

- Allen, R.E. (Ed.).** 1990. *The Concise Oxford Dictionary of Current English*. New York: Oxford University Press.
- Guralnik, D.B. (Ed.).** 1981. *Webster's New World Dictionary of the American Language*. Student Edition. Englewood Cliffs, New Jersey: Simon and Schuster.
- Marole, L.T.** 1954. *English–Venda Vocabulary*. Sibasa: Marole Book Depot.
- Wentzel, P.J. and T.W. Muloiwa.** 1982. *Improved Trilingual Dictionary: Venda–Afrikaans–English*. Pretoria: University of South Africa.

Other Sources

- Al-Kasimi, A.M.** 1983. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
- Baker, M.** 1992. *In Other Words: A Coursebook on Translation*. London: Routledge.
- Hammel, E.A.** 1966. *A Factor Theory for Aruta Kinship Terminology: Anthropological Records*, Volume 24. Berkeley/Los Angeles: California Press.

- Larson, M.L.** 1984. *Meaning-based Translation: A Guide to Cross-language Equivalence*. New York: University Press of America.
- Mandela, N.R.** 1995. *Long Walk to Freedom: The Autobiography of Nelson Mandela*. London: Routledge.
- Mbananga, N.** 2003. Multi-cultural and Multi-lingual Society: A Challenge for e-Health in South Africa. De Schryver, Gilles-Maurice (Ed.). 2003. *Sixth International TAMA Conference: Conference Proceedings*. Pretoria: (SF)2 Press.
- McHenry, R. (Ed.)**. 1992a. *The New Encyclopaedia Britannica. Volume 6*. Chicago: Encyclopaedia Britannica Inc.
- McHenry, R. (Ed.)**. 1992b. *New Encyclopaedia Britannica. Volume 18*. Chicago: Encyclopaedia Britannica Inc.
- Mönnig, H.O.** 1978. *The Pedi*. Pretoria: J.L. van Schaik.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 2002. Reversing an African-language Lexicon: The Northern Sotho Terminology and Orthography No. 4 as a Case in Point. *South African Journal of African Languages* 22(2): 161-185.
- Scheffler, H.W. and F.G. Lounsbury.** 1971. *A Study in Structural Semantics: The Siriono Kinship System*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Stone, L.** 2000. *Kinship and Gender: An Introduction*. Second Edition. Colorado: Westview.
- Van Warmelo, N.J.** 1904. *Kinship Terminology of the South African Bantu*. Pretoria: Government Printers.
- Zgusta, L.** 1971. *Manual of Lexicography*. The Hague/Paris: Mouton.

Polysemy and Homonymy: Challenges Relating to Lexical Entries in the *Sesotho sa Leboa– English Bilingual Dictionary**

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

Abstract: This article outlines the challenges relating to the lemmatization of the lexical items which are either polysemous or homonymous, as experienced during the compilation of the *Sesotho sa Leboa–English Bilingual Dictionary*. These problems can be ascribed to a lack of objectivity resulting from an inadequate knowledge regarding the etymological relationships of the meanings of some lexical items which are subsequently misclassified as either homonyms or polysemous words. This often causes improper lexical entries in dictionaries, i.e. polysemous words may be lemmatized as homonyms and vice versa.

To unambiguously distinguish between lexical items which are either homonyms or polysemous words, lexicographers should in this regard consider the criteria suggested by scholars:

- In accordance with the relatedness/unrelatedness criterion, the lexicographer will need to determine the extent to which the lexical items are related before entering them in the dictionary.
- The etymological criterion will help the lexicographer to determine the relatedness of the lexical items for inclusion in the dictionary according to their historical connection.

Using these criteria to lemmatize a lexical item properly in the dictionary, the lexicographer will be required to determine whether a lexical item can be regarded as one word with more than one meaning (a polysemous word), or as two different lexical items which have the same spelling (homonyms). To make this distinction, the lexicographer will need to research the historical background of the lexical item. Even though these criteria are important aids for distinguishing polysemy and homonymy, the problem of subjectivity is not altogether removed since the criteria do not specify the level or degree of relatedness/unrelatedness at which the lexical item(s) can be said to be polysemous or homonymous.

Keywords: SEMANTIC SHIFT, POLYSEMY, HOMONYMY, MEANING, SENSE RELATION, LEXICAL ITEM, ETYMOLOGY, METAPHOR, METAPHORIC SENSE, EUPHEMISM, LEMMATIZATION

* This article was presented as a paper 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.

Opsomming: Polisemie en homonimie: Uitdagings betreffende leksikale inskrywings in die *Sesotho sa Leboa–English Bilingual Dictionary*. Die artikel skets die uitdagings betreffende die lemmatisering van leksikale items wat óf poliseem óf homoniem is soos teëgekomp gedurende die samestelling van die *Sesotho sa Leboa–English Bilingual Dictionary*. Hierdie probleme kan toegeskryf word aan 'n gebrek aan objektiwiteit wat die gevolg is van 'n onvoldoende kennis aangaande die etimologiese verwantskappe van die betekenisse van sommige leksikale items wat gevolglik foutiewelik geklassifiseer word as óf homonieme óf poliseme woorde. Dit veroorsaak dikwels verkeerde leksikale inskrywings in woordeboeke, d.w.s. poliseme woorde mag as homonieme gelemmatiseer word en andersom.

Om ondubbelsinnig te onderskei tussen leksikale items wat óf homonieme óf poliseme woorde is, behoort leksikografe in dié verband die kriteria te oorweeg wat deur vakkundiges voorgestel is:

- In ooreenstemming met die verwantskaps-/nieverwantskapskriterium sal dit vir die leksikograaf nodig wees om te bepaal watter leksikale items verwant is voordat hulle in die woordeboek ingesluit word.
- Die etimologiese kriterium sal die leksikograaf help om die verwantskap van leksikale items vir insluiting in die woordeboek te bepaal volgens hul historiese samehang.

Deur hierdie kriteria te gebruik om 'n leksikale item korrek in die woordeboek te lemmatiseer, sal van die leksikograaf vereis word om te bepaal of 'n leksikale item beskou kan word as een woord met meer as een betekenis ('n poliseme woord), of as twee leksikale items wat dieselfde spelling het (homonieme). Om hierdie onderskeid te tref, sal die leksikograaf die historiese agtergrond van die leksikale item moet navors. Selfs al is hierdie kriteria belangrike hulpmiddels vir die onderskeiding van polisemie en homonimie word die probleem van subjektiwiteit nie heeltemal oorkom nie, omdat die kriteria nie die vlak of graad van verwantskap/nieverwantskap aandui waarby die leksikale item(s) as poliseem of homoniem beskou kan word nie.

Slutelwoorde: BETEKENISVERSKUIWING, POLISEMIE, HOMONIMIE, BETEKENIS, BETEKENISVERWANTSKAP, LEKSIKALE ITEM, ETIMOLOGIE, METAFOR, METAFORIESE BETEKENIS, EUFEMISME, LEMMATISERING

1. Introduction

It is not always possible to distinguish polysemous lexical items from homonyms, and whenever this distinction is made, subjectivity prevails. The lexicographer's knowledge of the etymological development of the lexical items is of vital importance. In fact, the lexicographer shows the distinction made by entering homonyms separately in the dictionary, i.e. as two or more different lexical items even though the lexical items have the same spelling and/or pronunciation, while a polysemous lexical item is entered as one lexical item with its definition showing all its multiple meanings, i.e. two or more meanings attached to one lexical item.

2. Defining polysemy and homonymy

Several linguists, semanticists and lexicographers define polysemy and homon-

ymy accurately in academic and scientific researches. Stern (1931), Ullman (1962), Lyons (1977), Leech (1981), Hurford and Heasley (1983), Taylor (1989), and Louwrens (1994) are among those scholars who researched these sense relations. According to Leech (1981: 228), homonyms are 'roughly two or more words having the same pronunciation and/or spelling, but different in meaning'. This definition of homonymy is in line with that of Macdonald (1977: 625) who defines a homonym as 'a word having the same sound and perhaps the same spelling as another, but a different meaning and origin'.

From these definitions it can be deduced that homonymy covers both written and spoken forms, but it is possible to have partial homonymy (or heteronymy), where the identity is within a single medium as in homography and homophony (Crystal 1991: 167). The *Encarta Concise English Dictionary* (2001) defines a homograph as 'a word that is spelt in the same way as one or more other words but is different in meaning' and a homophone as 'a word that is pronounced in the same way as one or more other words but is different in meaning and sometimes spelling'.

In Sesotho sa Leboa there are, however, also tonal changes which need to be considered when describing homonyms. If two or more words are spelt the same, the tonal patterns must also be taken into account: heterotonal homonyms have different tonal patterns (Louwrens 1994: 75), while homotonal homonyms have the same tonal pattern (Louwrens 1994: 76).

The following are examples of two groups of words spelt the same in Sesotho sa Leboa:

fo la	cool down
fo la	line up (to queue)
fo la	smoke (a pipe, cigarette, etc.)
fo la	samp, crushed maize
mo golo	salary
mo golo	throat

The words in the first group of four, however, do not have the same pronunciation: the first two are pronounced with the mid-low vowel /ɔ/, therefore /fɔla/, while the last two are pronounced with the mid-high vowel /o/, therefore /fola/. But when the tonal patterns of these two pairs of words are considered, the following variations emerge:

- [fɔla]: *fólà* 'cool down' (high-low tone) and *fólà* 'line up' (low-low tone),
and
[fola]: *fólà* 'smoke' (low-low tone) and *fólà* 'samp, crushed maize' (high-low tone).

The words in this group are therefore heterotonal homonyms.

When the pronunciation and tonal patterns in the second group of two are investigated, it appears that they are the same in each case. Both words have the pronunciation /moyɔlɔ/ and the tonal pattern mògòlò (low-low-low). The

words in this group are therefore homotonal homonyms.

According to Taylor (1989: 103), homonymy occurs when unrelated meanings attach to the same phonological form. He distinguishes two ways in which homonymy is usually brought about, i.e.

- (a) firstly, related meanings of a once polysemous word have drifted so far apart that there is no perceived relationship between them, and
- (b) alternatively, unrelated words which were once phonologically distinct have been subject to the 'blind' operation of sound change, and in the course of time have become phonologically identical.

Hurford and Heasley (1983: 123) say that a case of polysemy is one where 'a word has several very closely related senses'. And Mojela (1991: 31) also explains it as a case where 'one word may have a set of more than one different but related meanings'. According to Taylor (1989: 99), polysemy is the association of two or more related senses with a single linguistic form. Illustrating it with an example, he says: 'The word "bird" can refer to many different kinds of creature — robins, penguins, ostriches, etc. These different kinds of creature are members of the category in virtue of similarity to a single prototype representation.'

3. Lexicographic challenges

The major challenge facing lexicographers is the making of a clear and objective distinction between polysemous and homonymous lexical items which are entered as lemmata in their dictionaries. Such a distinction is not always possible with all lexical items in a language. Lexicographers with little knowledge of the etymology of the lexical items in their languages will have problems in entering polysemous and homonymous lexical items in their dictionaries, since they will not know the extent to which the lexical items are related. In this regard, Lyons (1977: 550) says the following: 'The difference between homonymy and polysemy is easier to explain in general terms than it is in terms of objective and operationally satisfactory criteria.'

Lyons' statement shows that the distinction between polysemy and homonymy is more complex than is generally perceived. He gives two criteria to simplify the complexities of identifying polysemy from homonymy, i.e.

- the etymological criterion, and
- the relatedness/unrelatedness criterion.

4. The etymological and relatedness/unrelatedness criteria

With these criteria, the lexicographer's knowledge of the etymological development (or history) of lexical items will help in identifying their origin and

their relationships for entering them in the dictionary as homonyms or polysemous lexical items. The etymological criterion will assist the lexicographer in establishing whether the lexical items developed or originated from what Taylor (1989: 103) regards as 'related meanings of a once polysemous word which have drifted apart'. The origin and relationship between the following groups of lexical items can be considered:

tshela	jump
tshela	six
tšhela	pour (liquid, sand, etc. into a container)
tšhela	pay tax
šupa	seven
šupa	(to) point

The etymological analysis of the lexical item **tshela** 'six' will show that the word originated from **tshela** 'jump'. Counting in the Sotho tradition is done with the help of the fingers starting with the fingers of the one hand, and when the fifth finger is reached, 'jumping' to the other hand to find the sixth finger, in this way giving the number **tshela** 'six' its name. Obviously, the lexicographer with little (or no) knowledge of the etymology of this lexical item will enter **tshela** twice as two unrelated homonymous lexical items, i.e.

tshela	jump
tshela	six

while the knowledgeable lexicographer will enter it as one polysemous lexical item with two related meanings, i.e.

tshela	jump
	six

The etymological investigation of the lexical items **tšhela** 'pour (liquid, sand, etc.)' and **tšhela** 'pay tax' reveals that the second lexical item originated from the first one. The pouring of money into the tax-collector's coffer is metaphorically compared to the pouring of liquid, sand, etc. into a container. In this way, the two meanings become related, so that **tšhela** should be regarded as a polysemous lexical item.

The etymological relationship between the meanings of **šupa** 'seven' and **šupa** 'to point' can be explained by the fact that the number seven got its name from the seventh finger, which is used for pointing. The seventh finger or the 'pointing finger' is also referred to as **tšhupa baloyi**, which means the finger used to point out witches/wizards, i.e. to indicate who performed witchcraft/wizardry. Therefore, this relationship between **šupa** 'seven' and **šupa** 'to point' gives the lexicographer a reason to enter **šupa** as a polysemous word instead of two homonymous words.

5. Subjectivity associated with polysemy and homonymy

As a result of identification problems, objectivity will not always be possible for lexicographers dealing with polysemous and homonymous lexical items. Consequently, the lexical items which one lexicographer might regard as polysemous, the other one might see as homonymous. As Lyons (1977: 550) puts it: 'Relatedness of meaning is a matter of degree. Those lexical items which one person might regard to be semantically related to a certain degree, the other person might see them to be very far apart.'

This is especially the case when the lexicographer must decide about the degree to which the 'related meanings of a once polysemous word have drifted ... apart' so that there is 'no perceived relationship between them' (Taylor 1989: 103). The decision on how far words have drifted apart is mainly based on subjectivity. The following groups of lexical items can be discussed as examples:

tšhoša	frighten, terrorize
tšhoša	weapon (which frightens, terrorizes)
makarapa	helmets
makarapa	migrant workers from the Witwatersrand
dirapa	gardens
dirapa	game reserves
dirapa	graveyards

Lexicographers will not to the same degree perceive the shift which has occurred in the meanings of the 'once polysemous lexical item' **tšhoša**, i.e. 'to frighten, terrorize' and 'a weapon (which frightens, terrorizes)'. While one will regard these meanings to be related because **tšhoša** for 'a frightening weapon' originated from the metaphoric use of **tšhoša** 'frighten', the other will consider the two meanings to have drifted 'so far' apart that they should be regarded as homonyms, because 'weapon', which is a noun, cannot be associated with a verb 'frighten'.

Makarapa 'helmets' (singular: **lekarapa**) are safety hats worn by workers in the mines. Most of the men from the rural areas worked as migrant labourers in the Witwatersrand area. In those days, it was prestigious to work in the mines, returning during Christmas holidays (usually with a lot of money). That is why most of the returning migrant workers wore their helmets (**makarapa**) to show that they come from the mines. As a result, the name for the migrant worker's helmet **lekarapa** (plural: **makarapa**) metaphorically came to refer to the migrant worker himself.

The same semantic shift from what may be regarded as a polysemous relationship to a homonymous relationship is found in the lexical item **dirapa** 'gardens', 'graveyards', and 'game reserves'. The original meaning of **dirapa** is 'gardens'. As a result of euphemism and metaphoric associations, **dirapa** came

to refer also to 'graveyards' and 'game reserves'. The meaning shift is so drastic that only the researcher with a knowledge of the etymology of this lexical item would be able to trace the relationship between the three meanings of **dirapa**.

6. Conclusion

The main objectives of this article can be summarized by the answers to the following questions:

- Are lexicographers aware of the complications pertaining to the identification and the ultimate distinction between polysemy and homonymy?
- Do lexicographers realize that it is very difficult, if not impossible, to be objective in distinguishing between homonymy and polysemy?
- Are lexicographers aware of the importance of having sufficient knowledge of the etymological history of lexical items before distinguishing between polysemous and homonymous lexical items?
- Do lexicographers know that relatedness of meaning is only measured in degrees, and that the degree to which lexical items are related depends on the individual researcher's subjective conclusion?

A knowledge of the suggested criteria for identifying polysemy and homonymy is important in making a distinction between these two sense relations, but this does not always solve the question of subjectivity in distinguishing between polysemy and homonymy, since the criteria are always associated with the semantic relationship of the lexical items. Furthermore, it is obvious that the lexicographer will be faced with numerous lexical items for which it will not always be possible to trace the etymological history.

References

- Crystal, D. 1991. *A Dictionary of Linguistics and Phonetics*. Third Edition. Oxford/Cambridge, Mass.: Blackwell Publishers.
- Hurford, J.R. and B. Heasley. 1983. *Semantics: A Coursebook*. Cambridge: Cambridge University Press.
- Leech, G. 1981. *Semantics: The Study of Meaning*. Harmondsworth: Penguin Books.
- Louwrens, L.J. 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika.
- Lyons, J. 1977. *Semantics, Volumes 1 and 2*. Cambridge: Cambridge University Press.
- Macdonald, A.M. 1977. *Chambers Twentieth Century Dictionary*. Bath: Pitman Press.
- Mojela, V.M. 1991. *Semantic Changes Accompanying Loan-words in the Northern Sotho Lexicon*. Unpublished M.A. Dissertation. Pretoria: Vista University.
- Rooney, K. (Ed.-in-chief). 2001. *Encarta Concise English Dictionary*. London: Bloomsbury.
- Stern, G. 1931. *Meaning and Change of Meaning*. Gothenburg: Goteborgs Hogskolas Arrskrift.
- Taylor, J.R. 1989. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.
- Ullman, S. 1962. *Semantics: An Introduction to the Science of Meaning*. Oxford: Basil Blackwell.

Albert Busch. *Diskurslexikologie und Sprachgeschichte der Computertechnologie.* 2004, XIV + 478 pp. ISBN 3-484-31252-1 (Pb.). Reihe Germanistische Linguistik 252. Tübingen: Max Niemeyer. Price: €138.

Diese umfangreiche Monographie stellt die 2003 an der Philosophischen Fakultät der Göttinger Universität vorgelegte Habilitationsschrift des Verfassers dar. Sie besteht aus zehn Kapiteln, Resümee, Literatur- und Quellenverzeichnis.

Die Hauptfrage der Untersuchung lautet: „Wie ist die Computertechnologie als Leittechnologie der Gegenwart gemeinsprachlich diskursiv verarbeitet worden und welche sprachwissenschaftliche Konzeption ermöglicht es, diesen sprach- und diskursgeschichtlichen Prozess so zu analysieren, dass insbesondere die sprachliche Arbeitsteilung zwischen Experten und Laien als diskurskonstitutive Dimension sichtbar wird?“ (S. 1). Die empirische Grundlage der Untersuchung bilden 136 Parlamentsreden (von 1963 bis 2000), 216 Artikel aus dem Magazin *Stern* (1968 bis 2000) sowie die Online-Korpora des Instituts für Deutsche Sprache in Mannheim und das Leipziger Online-Wortschatzlexikon. Im 1. Kapitel weist der Verfasser u.a. darauf hin, dass der gemeinsprachliche Computerdiskurs zugleich ein fachexterner Diskurs ist; für solche fachexterne Diskurse „ist eine vertikale Differenzierung zwischen fachsprachlichen und gemeinsprachlichen Signifikanten- und Bedeutungsfeldern charakteristisch“ (S. 3). Gegenstand des Kapitel 2 sind Grundbegriffe einer vertikalitätstheoretischen Diskurslexikologie. Unter Vertikalität wird „die sprachliche Variation zwischen Experten und Laien eines Faches“ (S. 11) verstanden. Ein Diskurs ist „ein thematisch definierter öffentlicher Kommunikations- und Wissensraum“ (S. 12); er wird durch sprachliche, vertikal organisierte Diskursbehandlungen konstituiert. Die wichtigste Voraussetzung für die Diskurskommunikation zwischen Experten und Laien ist Öffentlichkeit.

Im 3. Kapitel diskutiert der Verfasser die Diskurstheorie von Michel Foucault und im nächsten erweitert er sie durch die Interdiskurs- und Kollektivsymboltheorie von Jürgen Link. Unter Kollektivsymbolik versteht Link „den Gesamtbereich der Symbolik, 'Bildlichkeit', Metaphorik, der anschaulichen Stereotypen und Klischees“ (S. 96). Der Verfasser betont die Integrations- und Übersetzungskraft von Kollektivsymbolen — sie können beim vertikalen Wissenstransfer zwischen Experten und Laien das Expertenwissen verständlich machen und veranschaulichen; sie ermöglichen andererseits Manipulation und Persuasion. In Kapitel 5 wird die von Siegfried Jäger vorgeschlagene Modellierung des Diskurses als „Fluß von Wissen durch die Zeit“ ausführlich dargestellt und dessen Brauchbarkeit für die vertikalitätstheoretische Diskurslexikologie kritisch diskutiert. Das 6. Kapitel behandelt das Modell der historischen Diskurssemantik von Dietrich Busse, der den Ansatz von Foucault weiterentwickelt, indem er die Handlungs-, Wort- und Textebene in die Analyse des Diskurses integriert.

Nach den korpuslinguistischen und methodologischen Präzisierungen zum Diskurs und Diskurskorpus in Kapitel 7 wird in dem nächsten Kapitel ein

Modell zur Analyse des Computerdiskurses vorgeschlagen. Der Verfasser konzentriert sich auf drei Fragenkomplexe:

- Verarbeitung der Computertechnologie im öffentlichen gemeinsprachlichen Diskurs (*Diskursprogression*).
- Die Entwicklung der Sprach- und Diskursgeschichte der Computertechnologie und ihre Abhängigkeit von der Technologiegeschichte (*Diskurspersuasion*).
- Lexikalische Veränderungen in der Gemeinsprache, die die Computertechnologie herbeigeführt hat (*Diskurswortschatz/Diskursvertikalität*).

Im Weiteren werden Instrumente zur Analyse der Diskursprogression behandelt; als Hauptquelle zur Erfassung der Thematisierungsfelder dient das zentrale Hilfsmittel der Parlamentsdokumentation — der Parlamentsthesaurus (PARTHES) (2001 enthielt er über 70 Wortfelder). Zur Analyse der lexikalisch basierten Diskursvertikalität wird „WordSmith-Tool“ von Oxford University Press verwendet; das *Wortfamilienwörterbuch der deutschen Gegenwartssprache* dient zur Wahl der computerrelevanten Basislexeme. Bei der Untersuchung der Diskurspersuasion wird den Technologietopi besondere Aufmerksamkeit geschenkt.

In Kapitel 9 werden das Diskurskorpus und die Spezifik der repräsentierten Bereiche der parlamentarischen Kommunikation und der Kommunikation der Populärpresse ausführlich diskutiert. Kapitel 10 enthält eine umfangreiche Analyse zur gemeinsprachlichen Verarbeitung der Computertechnologie; der Verfasser unterscheidet sechs Phasen der Technologieentwicklung und der Diskursentwicklung. In jeder Phase werden die Diskursprogression, Diskurspersuasion sowie der Diskurswortschatz und seine Vertikalität breit beschrieben und mit vielen Beispielen belegt.

Eine kurze Zusammenfassung rundet diese tiefgründige und sowohl theoretisch als auch empirisch gut abgesicherte Untersuchung ab; die Monographie ist als ein wichtiger Beitrag zur Sprachgeschichte als Kulturgeschichte und zur Diskurslexikologie zu bewerten.

Andrzej Katny
Institut für Germanistik
Universität Gdansk
Gdansk
Polen
(*akatny@wp.eu*)

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume I: 1567–1785.* 2004, xii + 259 pp. ISBN 0 19 925470 2 (Hb.). Oxford: Oxford University Press. Price: £45.

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume II: 1785–1858.* 2004, xiv + 338 pp. ISBN 0 19 925470 2 (Hb.). Oxford: Oxford University Press. Price: £45.

These two beautifully printed and presented volumes document a complete history of dictionaries of English cant and slang from 1567 to 1858. Coleman investigates the wealth of dictionaries and glossaries of cant and slang that are, as she puts it, 'far more often used than studied' but which, nonetheless, form an essential part of the 'tangled tradition' (I: 1) of early lexicographical study. She goes beyond just documenting this history, though, and the range of her work is wider than that implied by the title. She investigates the social history behind these dictionaries, the criminal (auto)biographies that gave rise to some of them, how these dictionaries in their turn were used in some literary texts from *Beggar's Opera* to *Guy Mannering* and how they contributed to a culture that became obsessively interested in crime. In the process Coleman turns what could have been dull chronology into a lively and intellectually-engaging account.

Coleman begins her documentation with a contextual history of criminality, including such matters as bridewells, Tyburn tickets, belly pleas, benefit of clergy, transporation, hulks and other standard topics in the history of crime. What is made clear from this history is that the wandering poor, from vagrants to runaway apprentices, were an object of great fear to the respectable citizenry and numerous laws were passed to restrict the movements of the poor across the countryside. Thomas Harman's *Caveat or Warening for Commen Cursetors* first appeared in 1567, just after changes to the parish poor laws had divided the poor into 'deserving' and 'undeserving' and made vagrancy both more common and more harshly punished. Appended to Harman's work was a brief glossary of beggars' language that begins the *History of Cant and Slang Dictionaries*. As a Justice of the Peace, Harman was concerned to enable householders to identify and then to understand the nefarious practices of the gypsy and the vagabond in order to be able to protect their properties. As Coleman wittily puts it, 'For Harman, the vagabond is lazy, parasitic, thieving, and deceitful. His was the progressive view' (I: 21). 'The cant dictionary, then,' argues Coleman, 'began its life as a practical measure for the suppression of crime and vagrancy: as a defensive tool for the magistrate and the householder' (I: 22).

Volume I, 1567–1785, deals with cant dictionaries, their origins and their influence on both language and literature. There were well over a hundred cant dictionaries and word lists published in this period so this is an ambitious undertaking. 'Cant' in this context means the secret language used by beggars and criminals to conceal their activities from their victims or the authorities. Its

very nature as an oral, secret and constantly-changing language would naturally tend to make its documentation, both by contemporaries in cant dictionaries, and by Coleman in her investigation of these texts, a very tricky enterprise, and, not surprisingly, etymology is a lexicographical feature often absent from these dictionaries.

Some of these terms, though they started life as cant, have, by whatever process (and their documentation in dictionaries of cant may well have been part of that process), become part of the mainstream language and could no longer be called cant, such as *to filch*, to steal. Coleman is fully aware of the potential unreliability of her sources and raises the (ultimately unanswerable) question throughout this history about how accurate these cant dictionaries are. She turns the focus of her attention from the empirical historical question of what these dictionaries tell us about the language of the underworld in this period to an investigation of what these dictionaries can tell us about contemporary attitudes to crime and its practices. Most of the recorders of cant were not themselves part of the criminal underworld and some, like Harman, represented the authorities, so there is at least the possibility that the criminals they consulted were deliberately misleading them, or even having a joke at their expense. There were exceptions such as Thomas Dekker and S.R., who both claimed to have personal experience of the underworld, though S.R. uses Dekker's *Bellman of London* (1608) lists and Dekker in turn uses Harman as the basis of their own lists, thus conforming to the wide practice of plagiarism in early lexicography and at least raising a question mark against their claims of authenticity. However, Coleman does include some texts written by authors who were later hanged for their crimes.

Like the chapter on 'The Harman-lists', the next chapter shows the influence of central canting texts, in this case Richard Head's *The English Rogue* (1665) and *Canting Academy* (1673), on their numerous imitators. Here Coleman charts a publishing history for the English–Cant list from *Canting Academy* that is significantly different from the Cant–English list. The latter is found mainly in editions of Head's own works while the English–Cant list is found in a wide variety of publications. Coleman concludes that 'the wider circulation of the English–Cant list may demonstrate that these glossaries were more popularly used for producing mock-cant than for understanding the real thing' (I: 65). She demonstrates that Head's *English Rogue* and *Canting Academy* were sources for the cant list contained in the anonymous *History of the Lives and Actions of Jonathan Wild* (1725), which recounted the life of the notorious criminal and thief-catcher Jonathan Wild along with Jack Sheppard and Joseph Blake.

Coleman credits Head with being the first to include 'citations' in his cant dictionary, but this seems to me to be a confusion between examples and citations. Citations are quotations taken from other people's writing and quoted as an instance of some word in use, whereas examples are most often made up by the lexicographer himself to exemplify some word or usage. Coleman argues that two of Head's entries are 'citations', included 'long before Johnson made

illustrative citations a normal feature of mainstream dictionaries' (I: 52). The first of these is:

Earnest A Part

As tip me my Earnest Give me my part or share

This, however, seems to be clearly an example and Johnson, who also includes examples as well as illustrative quotations, uses the same formula "as, ..." preceding all of his examples.

The chapter on B.E.'s *New Dictionary of the Terms Ancient and Modern of the Canting Crew* (1698 or 1699) is the part of the story of the development of cant dictionaries that moves into the mainstream. It was the first to appear as an independent volume rather than attached to a larger work, the first to include features found in mainstream dictionaries and the largest by far of the early canting dictionaries at over 4 000 entries. B.E. used several sources including Coles's *English Dictionary* (1676) and, Coleman lucidly demonstrates, Thomas Shadwell's play, *The Squire of Alsatia* (1688). Coleman provides an admirably thorough analysis of B.E.'s *New Dictionary*, though, again, a confusion about citations leads to some puzzling statements, such as, 'Where citations are at all predictable in this way, it is tempting to assume that B.E. just made them up' (I: 79) and she argues that multiple 'citations' are 'lexicographically unnecessary' (I: 78). It seems to me that B.E. is fulfilling the promise on the title page of including 'Proverbs, Phrases, Figurative Speeches, &c.' and in this he is also followed by Johnson along with many others.

The colourful lives of some of the people in Coleman's source material is part of the attraction of these volumes. One such is Bamfylde-Moore Carew who 'ran away to join the Gypsies and was elected their king, but was subsequently convicted of vagrancy and sentenced to transportation to Maryland' (I: 127), from where he escaped dressed as a Quaker and later avoided imprisonment by faking smallpox.

The dominant text of Volume II, which covers the period 1785–1858, is Francis Grose's *Classical Dictionary of the Vulgar Tongue* and the main subject of this volume is slang rather than criminal cant. Coleman also extends her history to cover the U.S.A. and Australia. Whereas the earlier canting texts were mainly anxious about rural crime among floating populations of gypsies, vagrants and beggars, later texts tend to concern themselves with urban crime. Coleman carefully distinguishes colloquial language, dialect terms, slang, jargon, cant and flash (fashionable slang of the demi-monde) at the outset of this volume. She estimates that 'by the 1780s, editions of cant and slang word-lists were being published at the rate of two each year' (II: 7), which seems an astonishing output for works on non-standard English. Coleman suggests that social distinctions in language may have arisen as a consequence of industrialization which mixed old money with new money, and a desire to document and preserve rural dialects may have been one of the consequences of rural migration. Again, Coleman includes a consideration of literature influenced by cant

or slang, such as *Oliver Twist*, and the way in which literature provided source material for slang dictionaries, such as Grose's use of Swift's *Polite Conversation*.

One problem throughout these volumes is an unfortunate one of timing. The OED is on the process of being fully revised and edited but Coleman has been forced back frequently on the unrevised version for comparison, commentary on its coverage of these early dictionaries and for antedatings, when these remarks will surely be rendered obsolescent by the appearance of the new OED.

One of the most fascinating parts of this work, though, is the word indexes, which are taken from the various cant and slang dictionaries but form, in effect, a sort of cant/slang dictionary in themselves:

cabal "a group of conspirators"
cabbage "the part of a deer's head where the horns grow"
cacafuego "a blustering man"
cackling-cheat *also* cackling chete "a chicken"
cackling-fart *also (erron)* crackling-fart "an egg"
cad "an untrustworthy man"
cadater "a beggar"

Included in an appendix in each volume are analytical tables of subject matter and lexicographical features of the main dictionaries covered. An additional appendix in Volume I contains canting songs. The bibliographies are followed by word and subject indexes.

This is a learned, admirably thorough and lively history of lexicography which considers dictionaries of cant and slang that are, as Coleman rightly says, more often used than studied. It is more than a linguistic and lexicographical study, though, since it includes a great deal of literature as well.

Anne McDermott
University of Birmingham
Birmingham
United Kingdom
(*a.c.mcdermott@bham.ac.uk*)

Christa Kilian-Hatz. *Khwe Dictionary*. With a Supplement on *Khwe Place-names of West Caprivi* by Matthias Brenzinger. 2003, x + 431 pp. ISBN 3-89645-083-2. Namibian African Studies 7. Cologne: Rüdiger Köppe Verlag. Price: €52.80.

Before an assessment of this dictionary is made, it is important to place the Khwe language in its historical and social context. Khwe dam (language of the Khwe), as its speakers would prefer to call it, is a Khoisan language belonging to the Southern African Central-Khoisan under the Khoe cluster (Güldemann and Vossen 2000). It is spoken in Northern Botswana and in the Western Caprivi Strip of Namibia in the Kavango region. As the compiler correctly observes, there are about 10 000 Khwe-speaking people who live in the Caprivi Strip of North-Western Namibia, Angola, Botswana and Zambia, and at Schmidtsdrift, South Africa. As can be expected from a speech community such as the Khwe, who prefer living in small communities, there are dialects within the Khwe dam. In Botswana, the notable varieties are the Buga (Sand Khwe) and the ||Ani (River Khwe), which exhibit discernable linguistic differences at the lexical and grammatical levels. However, linguists agree, as Kilian-Hatz also indicates, that the Khwe dam dialectal varieties are definitely related and mutually fairly intelligible (cf. Güldemann and Vossen 2000). As a language that has not been elaborately codified, it is difficult to speak about a standard Khwe dam form. Many other social and linguistic activities need to be carried out to resolve the question of standard Khwe dam (cf. WIMSA 2001).

As also reported in the introduction to the dictionary, the liberation war in Namibia caused some serious disturbances to the Khwe speakers' social ecology and geographical distribution. Some speakers have consequently been resettled in the Northern Cape Province of South Africa (Takada 2007). Yet this relatively recent social turmoil that affected them is just another of the historical occurrences in their precarious and necessitous lives. Within the Kavango region of Namibia and the Okavango delta of Northern Botswana, the Khwe people, as most Khoisan peoples elsewhere, have always come into contact with other ethnic groups, who often dominated them socially, economically and linguistically. There has therefore been some important lexical borrowings from and even a language shift towards influential regional languages. Also, there is very little that suggests the current language use policies in the countries in which they are found would change the situation for the better. The Khwe are marginalized and subjected to the ethnolinguistic assimilations of the more populous and better organized ethnic groups surrounding them.

It is against this background of a historical and sociolinguistic marginalized speech community that the importance of a lexicographical work such as this Khwe dictionary can be assessed. Recording a language in dictionary form for an endangered language such as Khwe is a milestone in the study and preservation of the language. The compilation of a dictionary, as also an act of language codification, provides a source for important linguistic data and indige-

nous knowledge systems, and as such is critical to all future research on the language. As indicated in the dictionary presentation, this compilation contains about 5 700 entries consisting of lexical as well as grammatical morphemes. In terms of African language standards, this compilation corresponds to a lexical stock that allows an ethnolinguistic speech community to sufficiently express its knowledge systems. Also, as a bilingual dictionary, the limitations are understandable. A monolingual dictionary would help regenerate more linguistic (lexical and grammatical) stock.

Furthermore, the introduction provides essential information on the language dealt with in this dictionary. It contains a map, which helps to precisely locate the speakers' territory. The Khwe dialects are listed as ||Xo-Khwe (lit. Dry Land Khwe, a dialect of the Khwe living mainly in the East Caprivi), ||Xom-Khwe (lit. River Khwe, the dialect of the Khwe living mainly on the Kavango River in West Caprivi), Búma-Khwe (lit. North Khwe, the dialect of the Khwe living in the Caprivi and in Botswana, but originating from Angola) and Búga-Khwe (lit. Bush Khwe, the dialect of the Khwe living in Botswana). It is important to note that this language is in transition and that there is evidence of dialectal convergence. There are also lexical borrowings (loan-words) from other languages such as Mbukushu, Afrikaans and English. This situation has some consequences for the use and variation of words within Khwe dialectology.

The dictionary consists of preliminary remarks and acknowledgements, the table of contents, abbreviations and symbols, an introduction, the Khwe-English dictionary and the English-Khwe index, the references section, and annexes consisting of Khwe proper names and Khwe place-names. The table of contents gives a synoptic view of the dictionary layout, important for providing practical guidance to the use of the dictionary. The table of abbreviations and symbols is elaborate, giving essential information on the abbreviations of the grammatical categories and the linguistic or dialectal sources of some lexical entries in the dictionary. These abbreviations and symbols reflect a language that is grammatically rich and the comprehensive analysis provided in the compilation is linguistically rewarding. For a language such as Khwe, this information facilitates the use of the dictionary for different language activities.

The format of presentation is therefore helpful in view of the language that is being treated lexicographically. The preliminary remarks, as brief as they may seem, are informative as they provide research data sources for the dictionary. The extensive research by the renowned linguist Oswin Kohler is, for instance, acknowledged. Kohler, who produced major works on Khoisan, is credited for the current understanding of Khoisan classification (cf. Güldemann and Vossen 2000; Kohler 1981; Legère 1998; Haacke and Elderkin 1997). The botanical data also comes from a well-known Khoisan researcher Mathias Schaldt (cf. Schaldt 2000). Several contributions on material culture by Gertrud Boden and Stefanie Michels are furthermore acknowledged. The acknowledgements also recognize the contribution of ethnic speakers, namely David Naudé

and Bothas Marinda Xunudao. The collaboration by native speakers ensures the reliability and representativeness of the dictionary data. The "insightful comments and invaluable help" by Wilfred Haacke is furthermore recognized. Haacke, a Khoisan lexicographer (Haacke 1998) himself, based at the University of Namibia, is crucial in assuring the quality of any dictionary publication on the Khoisan languages of Namibia.

The presentation of the orthographic conventions cannot be underestimated. While they are essentially adapted from Kohler, the compiler indicates that some practical enhancements have been introduced to make the dictionary user-friendly. There is further information that the speakers' own writing practices have been taken into account. For a dictionary which targets a minority and marginalized language, it is important that the community of speakers should not feel completely alienated by linguists' academic research works. It is this practical aspect of the dictionary that will remain a constructive contribution to the development of the language. The orthographic presentation gives an elaborate discussion of the writing symbols. Phonemically, as it can be observed, the Khwe language has an extensive system and an elaborate list of representative letter symbols, more than that of the Latin alphabet. If the order of entries is anything to go by, there are 55 alphabetized entries. The four fundamental clicks (dental, alveolar (retroflex), palatal and lateral) have phonemic accompaniments which necessitate the addition of other letter symbols to cater for voicing, aspiration, ejection, velarization, uvularization, nasalization, and glottalization. The other non-click consonant symbols numbering about 16 are also carefully described and approximate English equivalences indicated. This is most helpful in view of the readership of the dictionary, who may not necessarily have the phonological knowledge of the language. If the symbol combinations (bigraphs and trigraphs) account for most of the elaborate entry listing, it should be understandable that the language has a complex phonetic and phonological structure.

As the author pertinently indicates, Khwe is a tone language and its tonology also forms the core of the lexicon, the word morphology and the sentence structure (or syntax). Tone is marked on vowels and nasal consonants, which are the tone-bearing segments. The author indicates eight tones, consisting of three lexical tones with the practical marking of high (á), mid (ã), and low (à). So critical is the role of tone that information of verb tone classes is provided as they have a bearing on the grammar of the language and its derivations. Being linguistically researched, the dictionary takes other conventions of linguistic importance into account, such as morpheme boundary marking by a hyphen. However, as the author states, this is for purely lexicographical purposes and not part of the orthographic conventions.

Part two consists of the Khwe–English dictionary which constitutes the greatest part of the book, indicating that the target language is Khwe. The Khwe–English entries are elaborate, providing headwords, grammatical information (category, derivation, morphology), and definitions. The definitions are

brief, giving short descriptions or lexical or scientific (botanical, technical) equivalences. Khwe sentences are also used to enhance definitions and provide contextual meaning. As it grammatically contextualizes each headword, this approach is vital, especially for a language that needs to be preserved and utilized by the speakers in their dialectal variations.

Part three of the dictionary consists of the English–Khwe index in which the Khwe equivalences for the English and scientific words are given. The only information given on these Khwe equivalences is the grammatical categories and the semantic fields of some words. A commendable feature is that every Khwe word in the index also appears in the Khwe–English part of the dictionary, allowing for easy reference to and elaboration of its definition and other uses. The index therefore supports the Khwe–English dictionary, focusing deliberately on the study and presentation of the Khwe language.

The dictionary and the index are followed by an appendix on proper names. This brief onomastic presentation furnishes the type of name (first, clan, family or nickname), the grammatical source of the name (verb, morphology and source), socio-anthropological information (social events, situations, and conditions in naming) and the gender (the author states that as a rule there is no gender distinction in naming). Words in this section are not systematically referred to in the dictionary sections (i.e. sections two and three). This is not quite justifiable, as these categories of words are part and parcel of the composite linguistic and indigenous knowledge base of a speech community. Although it might be useful to have all the proper names together in one section, there should not really be any separation of words belonging to the same language.

The last part of the dictionary is the supplement on Khwe place-names in West Caprivi compiled by Matthias Brenzinger. The introduction to this section justifies the giving of this linguistic and ethnographical information on toponyms and their role in the cosmology of the Khwe people. Brenzinger indicates that there are 371 place-names collected from different areas where the Khwe people had settled. This is a well-researched data compilation solidly placing Khwe naming practices within a wider social and anthropological perspective. Indeed, names, as it is evident from the presentation, are not just limited to geographical place references, but to all the life experiences of the Khwe people. There are linguistic, social and historical values in the names given to places. In his vivid discussion, Brenzinger demonstrates that for languages, and especially for Khwe, there is much that can be derived from names, and indeed place-names. They map out the territory of the ethnic community's movements, socio-economic and historical, and the justification for their territorial attachment thereto. This section indeed gives proof of a profound onomastic study as well as a theoretical analysis of names. It will be valuable to onomatologists, toponymists and researchers in the related fields of the cultural and linguistic aspects of names.

To summarize: In the first instance, this Khwe dictionary and the supple-

ment on place-names provide an interesting compilation for a first edition of a reference work on a specific Khoisan language. The determining of the lemmas, the defining of the headwords and the treatment of the semantic domains make the dictionary quite practical for native speakers, for Khwe learners, and for linguists and grammarians. The lexicographical awareness in the handling of the Khwe–English entries greatly enhances the semantic value of the headwords and their application to different domains. This dictionary constitutes valuable work not only on the vocabulary of the Khwe language but also on the whole ethnolinguistics and ethnoculture of the Khwe. Essentially it is minimally bilingual. The emphasis is on Khwe, and the English facilitates its linguistic and lexicographical accessibility to other users than Khwe speakers. The English also enhances the functionality of the dictionary which will indeed facilitate comparative Khoisan studies. The dictionary will provide a resource for the speakers and a source of linguistic analysis for researchers.

In the second instance, as there is no comprehensive grammar published on the Khwe language yet, a grammatical sketch of Khwe would have been greatly appreciated. For the community and any other user, the availability of such publications that are the result of difficult research seldomly undertaken, remains the only available codified material for a considerable time. Very often, language teachers in such communities have nothing else to resort to in their endeavour to contribute to the promotion of the language through literacy. This comment, however, should not be construed to question the dictionary which has its own lexicographical purpose, and its compilers their own task, which is to record the lexicon of the language. The only wish that can be expressed is that there will be some follow-ups in other areas of Khwe linguistics to complement this dictionary.

As indicated in the introductory matter, the Khwe dictionary has benefited from various research and field-work and contributions from other researchers and field-workers. However, the compiler Christa Kilian-Hatz has demonstrated the assiduity of a field linguist. It has creatively combined an eclectic mix of lexical sources. This open-minded approach also accounts for the inclusion of the section on place-names by Brenzinger. All aspects cohere to make a good reference work. Even though it is bilingual, the focus is on the Khwe language. It will therefore be understandable why the English–Khwe section is more of an index than an elaborate lexicographical presentation. The importance of a lexicographical work not only lies in the practicality and functionality of its usage, but also in the forming and building of a language knowledge base — the lexicon. Even though this compilation may not be exhaustive in terms of collecting the complete Khwe vocabulary, it establishes a firm foundation for the lexicography of the language. It will serve as a research text for field linguists and other language analysts. Indeed, it will be a lexical companion for the speakers and a necessary text for the learner of the Khwe language. This dictionary cannot be compared with Haacke and Eisib's *Khoekhoegowab Dictionary* (2002) (cf. Haacke 1998), which has benefited from exten-

sive literary and linguistic sources spanning over a century. This is lacking for Khwe. Therefore, one can only be happy and grateful for such a pioneering work as this which Kilian-Hatz has effected. It will contribute much to the standardization of Khwe, which is crucial to language development (cf. Haacke 1989).

Having said all this, it will greatly enhance this dictionary if its next edition should not only expand on the vocabulary, but also merge the separate section on names into the main dictionary. In the true sense of a dictionary, words and names are lexically the same linguistic material. For practical purposes, looking for a word or name within the same alphabetic order of entries is not only economical but also desirable.

In conclusion, the strength of this dictionary lies in its judicious application of a lexicographical technique which varies the definitions with descriptions and translations of terms. Yet there are no excesses in the illustrations, lengthy semantic explanations being absent. These have a justifiable place in a monolingual dictionary. Indeed, this dictionary has made and will make an important contribution to the language development efforts of the Khwe community by adopting and justifying the practical orthography the speakers have accepted (cf. WIMSA 2001). There is therefore no need for conversion from phonetic to practical as the symbols used constitute the conventional orthography. This is a notable change as very often dictionaries in Khoisan languages have mainly been motivated by linguistic and not community considerations. As linguistic documents, dictionaries have remained on the shelves of libraries and academics' offices. It is therefore a truly appreciable development that this dictionary will be accessible for use by teachers, students and literacy material developers of Khwe. It should be emphasized that a dictionary plays a powerful role in language development; it canonizes the orthography and the word usage. As such, it is always desirable that the compilers should produce dictionaries that will stand the rigour of judgment and the test of time. Here the compiler has undoubtedly succeeded in taking the linguistic research on the Khwe language to a new level. This dictionary can be recommended to Khoisan linguists particularly, and, having a solid base for language development and promotion by the speakers themselves, to community language developers and promoters.

References

- Brenzinger, Matthias.** 2003. *Khwe Place-Names of West Caprivi*. Namibian African Studies 7. Cologne: Rüdiger Köppe Verlag.
- Güldemann, Tom and Rainer Vossen.** 2000. Khoisan. Heine, Bernd and Derek Nurse (Eds.). 2000. *African Languages: An Introduction*: 99-122. Cambridge: Cambridge University Press.
- Haacke, W.H.G.** 1989. Nama: Survival through Standardization? Fodor, I. and C. Hagege (Eds.). 1989. *Language Reform History and Future*: 397-429. Hamburg: Helmut Buske.

- Haacke, W.H.G. and Eliphaz Eiseb.** 2002. *A Khokhoegowab Dictionary with an English–Khoekhoegowab Index*. Windhoek, Namibia: Gamsberg Macmillan.
- Haacke, W.H.G. and E.F. Elderkin.** 1997. *Namibian Languages: Reports and Papers*. Cologne: Rüdiger Köppe Verlag.
- Haacke, Wilfrid.** 1998. A Khoekhoe Dictionary in the Making: Some Lexicographic Considerations. *Research in Khoisan Studies* 15: 35-64.
- Köhler, Oswin R.A.** 1981. Les langues Khoisan. Perrot, Jean (Ed.). 1981. *Les langues dans le monde ancien et moderne. Volume I: Les langues de l'Afrique subsaharienne — Pidgins et créoles*: 455-615. Paris: Éditions du C.N.R.S.
- Legère, Karsten.** 1998. Khoisan Traces in Kavango Languages. Schladt, Mathias (Ed.). 1998. *Language, Identity, and Conceptualization among Khoisan*: 197-216. Quellen zur Khoisan-Forschung 15. Cologne: Rüdiger Köppe Verlag.
- Schaldt, Mathias.** 2000. A Preliminary List of Kxoe Wild Plants. *Khoisan Forum* 14: 3-47.
- Takada, Akira.** 2007. Changing Locality and Ethnic Cohesion among San in Ohangwena, Namibia. Fujioka, Yuichiro and Masashi Iida (Eds.). 2007. *ASAFAS Special Paper 9*: 75-90. Kyoto: Graduate School of Asian and Africa Area Studies (ASAFAS).
- Working Group of Indigenous Minorities of Southern Africa (WIMSA).** 2001. *The Penduka Declaration on the Standardisation of Ju and Khoe Languages. 20–22 April 2001*. Windhoek, Namibia: Penduka Training Centre.

Andy Chebanne
Department of African Languages
University of Botswana
Gaborone
Botswana
(chebanne@mopipi.ub.bw)

Phillip Louw (Senior Redakteur/Senior Editor). *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. 2007, xii + 578 pp. ISBN 978 0 19 578742 9. Kaapstad/Cape Town: Oxford University Press Southern Africa (Pty) Ltd. Prys/Price: R99.95.

Die *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek* is as skoolwoordeboek 'n puik hulpbron vir leerders. Dit is egter meer as net 'n woordeboek, soos wat 'n oorsig dadelik toon.

Na die titel- en inhoudbladsy word 'n dubbelblad **Woordeboekkenmerke/Dictionary features**, aangebied wat in ander publikasies van dié aard as die *artikeluitleg* bekendstaan. Dit is kenmerkend 'n afbeelding van 'n dubbelblad uit die boek waarin die vernaamste kenmerke met pyltjies aangedui en verklaar word.

Die inleiding verstrekk onder meer die oorwegings by die saamstel van die werk. Een daarvan is om met die lemmas en uitdrukkings hedendaagse gebruikswyses en aangeleenthede te weerspieël. Hierdie voorwerk is sorgvuldig beplan.

Wat die hoofdeel betref: Soos wat al gebruik geword het, plaas voorbeeldsinne, en nie definisies nie, die lemmas in konteks; dié sinne word in albei dele van die woordeboek in albei tale aangebied (met 'n klompie hiate wat later bespreek word).

Tussen die Afrikaanse en Engelse deel is 'n **Studieafdeling/Study section** ingesluit wat die nut van die werk verhoog. Dit bevat oefeninge om 'n woordeboek te leer gebruik — 'n aspek wat min of geen aandag op skool kry nie; voorbeelde van 'n formele en informele e-posboodskap en -brief en 'n nuttige nuwigheid: 'n tabel met SMS-taal wat veral die taalgebruik op selfone van vandag se jeug weerspieël. Dit bestaan uit die kortpaaie wat geneem kan word om hierdie bydertydse manier van kommunikasie te bespoedig. Voorbeelde is: *n* vir die woord *en*, *8nee* vir *ag nee*, *kani* vir *kannie* en dergelike meer. (Die Engelse tabel bevat, soos verwag sou word, talle meer voorbeelde.)

Nog 'n nuutjie is die sogenaamde *emosikone*: simbole wat kan wys hoe die skrywer voel, byvoorbeeld "gelukkig" word uitgebeeld met 'n glimlag of lag op 'n gesiggie wat dwars gedraai is [:-)], [:-D], en "ongelukkig" met 'n gesiggie wat lyk asof dit huil [:-(].

Leestekens en basiese spelreëls gaan die Engelse kant van die woordeboek vooraf.

Hierdie veelheid van inligting word nog verder uitgebrei deur 'n **Verwysingsafdeling/Reference section** aan die einde wat afbeeldings van verskillende diere bevat, asook van groente en vrugte, die menslike liggaam en sport. 'n Kaart van Suid-Afrika met die provinsies, inligting oor Suid-Afrikaanse tale, 'n temareeks (maande, dae, datums, die sonnestelsel, openbare vakansiedae, seisoene, breuke en simbole), en afdelings oor getalle, en mate en gewigte sluit die werk af.

Oxford maak daarop aanspraak dat die woordeboek goedgekeur is vir gebruik in Suid-Afrikaanse skole en dit is maklik om te sien waarom — dit is 'n puik hulpmiddel vir leerders.

Die leksikografiese notas wat in hierdie resensie volg, moet gesien word met die teikenmark in gedagte. Daarmee word bedoel dat die leerder (en ook die deursnee-opvoeder) waarskynlik nie die leksikografiese hoogtepunte en 'ongelykhede' (om dit nie 'foute' te noem nie) by die praktiese gebruik van die boek sal raaksien nie. Laasgenoemde word hier dan ook aangebied vir oorweging by 'n volgende druk.

Eerstens enkele opmerkings rakende die andersins uitstekende voorwerk, "middelwerk" en agterwerk. Op bl. iv verskyn een van min setfoute wat teengekom is, naamlik *Nuttige frases and idioeme* in plaas van *en*, en op bl. 571 kom *Fase van onderwys* voor in plaas van *Fases* (AOO en VOO/GET en VET, in die Afrikaanse en Engelse uiteensettings onderskeidelik, word egter nêrens verklaar nie). Op bl. S15 word by homofone verwys na *fier/vier* en *peil/pyl*, maar ook **fyf/vyf*. Miskien is *fuij* in plaas van *fyf* bedoel, maar ook dié ontronding laat dit nie as homofoon kwalifiseer nie. Verder word op bl. viii *Woordeboek Werkboek* as twee woorde in plaas van een geskryf.

'n Inskrywing op bl. x (in Afrikaans) en ook so vertaal (op bl. xii) wat sekerlik 'n verskrywing is, is: 'Vir die meeste Engelse werkwoorde is die *verlede deelwoord* en die *teenwoordige deelwoord* dieselfde [my kursivering]. Dit is nie moontlik nie: die morfologie in die boek wys ook dat die bedoeling was dat die *eenvoudige verlede tyd* en die *verlede deelwoord* in verreweg die meeste gevalle dieselfde is, bv. *play (playing, played)*.

'n Ander lastigheid (wat ook in die lemmakorpus na vore kom) is die gebruik van *onderwyser(es)* (bl. S2) in plaas van *opvoeder*; dit is veral opvallend omdat die (tans korrekte en voorgeskrewe) *leerder* wel deurgaans gebruik word. *Opvoeder* is ook nie aan die Afrikaanse kant opgeneem nie. Dié nuwe benaminge kom uit artikel 1 (die woordoms krywingsartikel) van die Suid-Afrikaanse Skolewet, Wet no. 84 van 1996.

Nou na die makrostruktuur. Die keuse van lemmas (altd 'n netelige saak) is by skoolwoordeboeke 'n groter probleem as by algemene woordeboeke. In Oxford se geval is daar met groot oorleg gewerk en die resultaat vir beide Afrikaans en Engels is bevredigend. Die oorweging om hedendaagse woorde en uitdrukkings op te neem wat met bydertydse aangeleenthede verband hou, vind stewig neerslag in die lemmakorpus. Daar is ook versigtig en suinig te werk gegaan met samestellings, en dié wat ingesluit is, het 'n hoë gebruiksfrekwensie by die jonger geslag. Voorbeelde van Afrikaanse lemmas wat opgeneem is, is:

die voornaamwoorde **daaraan, daardeur, daarin, daarmee, daaronder, daarop, daarteen, waarby, waardeur, en waarin** wat 'n welkome bevestiging is dat die korrekte Afrikaanse manier nog nie deur die Engels beïnvloede **aan dit, *deur dit*, e.s.m. verdring is nie;

diens (motor) (Engels *service*) is reeds opgeneem in die *Afrikaanse woorde-lys en spelreëls* (AWS) en kan nou maar die ongebruiklike *versien* vervang;

dis (Engels *dish*) vir die puristiese *gereg* (ook AWS);

- fynbos** (omdat die Afrikaans ook in Engels gebruik word);
gemeen — *iets in gemeen hê*: ook reeds deur die Taalkommissie aanvaar naas *iets gemeen hê*;
jonk en **lank** (wat tot dusver in die AWS ontbreek het: die konsonantwysiging in verboë vorme soos *jonge* en *lange* word hiermee goed belig);
langtermyn- ('n nuttige koppeltekenlemma wat dié soort samestellings toelig, byvoorbeeld **langtermynprojek**);
lewenstandaard (sonder die *-ss-*, dit wil sê nie **lewenstandaard* nie, wat menige skrywer laat struikel);
muffin — aan die Afrikaanse kant opgeneem; by **kolwyntjie** word *cupcake* gegee, maar by **cup-cake** die foutiewe spelling *kolwynkie*;
ooit — wat die opname van 'n uitdrukking soos *die beste ooit* naas *die beste tot nog toe* moontlik maak; en
sigblad (Engels *spreadsheet*).

'n Hiaat wat opval is dat die wiskundeterm **parallelogram** (Afrikaans) opgeneem is met die voorbeeldsin: "n **Parallelogram** se teenoorstaande sye is parallel en ewe lank." Die lemma **parallel** ontbreek egter, wat die voorbeeldsin onduidelik maak. Aan die Engelse kant is **parallel** en **parallelogram** opgeneem, en die Afrikaanse voorbeeldsin by laasgenoemde is: '*n Parallelogram is 'n vorm met vier sye en die teenoorgestelde sye en hoeke gelyk* — wat die betekenis van die lemma dus verduidelik.

Die opmerkings vir Afrikaans geld ook die Engels. Een besonder puik geval is die hantering van **then**. Daar word duidelik onderskei tussen Afrikaanse toekomstige tyd *dan* en verlede tyd *toe* — 'n sjibbolet vir Engelse moedertaalsprekers wat Afrikaans aanleer.

Onder die letter Z is geen lemmas opgeneem nie; een wat nuttige leiding kon gee is **zero** omdat die Engelse *zero* 'n tyd lank as **sero* vertaal is. (Die wisseling *sero/zero* word egter wel op bl. 574 gegee.)

Daar is etlike gevalle waar die woordeboek nie "sluit" nie, dit wil sê waar lemmas in die een taal verskyn maar in die ander ontbreek. Dit is moontlik daaraan toe te skryf dat 'n frekwensietelling gebruik is en dat die sê nou maar gebruiklikste 6 000 woorde in Afrikaans nie noodwendig die gebruiklikste 6 000 woorde in Engels is nie, maar omdat dit nêrens vermeld word nie, bly dit vir die oningeligte dus 'n tekortkoming. Lukrake voorbeelde is Engels **ozone**, **parrot**, **pear**, **pea**, **peach**, **pineapple**, **placenta**, **plenty** en **rat** waarvan die vertalings nie aan die Afrikaanse kant verskyn nie; daarteenoor ontbreek byvoorbeeld Afrikaans **kontrasterend** se vertaling *contrasting* ook. (By hierdie lemmas is ook voorbeeldsinne nie opgeneem nie.)

Die AWS is ongelukkig in 'n aantal gevalle by die opname van lemmas nie geraadpleeg nie: **buurt** is opgeneem, die wisselvorm **buurte** nie; die koppeltekenlemma **chrono-** is opgeneem, **krono-** nie; **professioneel/professional** is slegs as byvoeglike naamwoord opgeneem, nie ook as selfstandige naamwoord nie. Daar is ook nie eenvormigheid in die morfologiepatroon van sommige ver-

lede deelwoorde wat as byvoeglike naamwoorde gebruik word en volgens die AWS met of sonder slot-**d** geskryf word nie, byvoorbeeld: **gebalanseer(d)** moet wees **gebalanseer(d)**, soos in die geval van **geïnteresseer(d)**, **gekleed(d)** en **uitgebrei(d)**; **gedifferensieer** moet wees **gedifferensieer(d)**. By **gespesialiseer** is die wisselvorm **gespesialiseer(d)** apart opgeneem, en so ook by **gestrem gestrem(d)**.

Dit is ook jammer dat **dagvaar** se afgeleide selfstandige naamwoord **dagvaarding** nie verskyn nie, aangesien die polemieë oor *dagvaarding*/**dagvaardiging* reeds 'n paar dekades lank duur.

Waar daar talle wisselvorme in Afrikaans is, is hulle in dié woordeboek tereg beperk. Een wat opvallend bydertyds is en reeds in die AWS verskyn, is die selfstandige naamwoord **gemiddeld** met die wisselvorm **gemiddelde**.

Algemeen gebruiklike afkortings (met die volvorm daarby) is opgeneem, byvoorbeeld **dr.**, **FM**, **ens.**, **DNS**, **Vigs**, **MIV**, **nm.**, **vm.**, **v.C.**, **n.C.**, **OTM**, **SMS**; dieselfde geld prefikse wat as koppeltekenlemmas verskyn, byvoorbeeld **pro-**, **self-**, **sub-**, **super-**, **teen-**, **twee-**.

Godsdienstige dae waarvan die name opgeneem is, verteenwoordig uiteenlopende religieuse stelsels: **Chanoeka**, **Rosh Hashana** (maar die heiligste dag vir die Jode en een van die gebruiklikste, **Jom Kippoer**, ontbreek); **Eid**, **Ramadan** (die AWS het slegs die spelling *Ramadaan*); **Diwali** (die AWS gee *Divali* en *Dipavali*).

Die mikrostruktuur is deeglik beplan, veral wat die volgende aspekte betref:

- (a) Morfologiese ingrepe soos *-e*, *-er*, *-ste* is nie gebruik nie; meervoude, verkleiningsvorme en trappe van vergelyking is voluit gegee, dus **kam**, **kamme** (Engels **comb**, **combs**); **gat**, **gaatjie** (ook gawe leiding oor die verdubbeling van die vokaal); **bly²**, **blye**, **blyer**, **die blyste**. Laasgenoemde vorm verteenwoordig 'n deurbraak in Afrikaanse woordeboeke. Die tradisionele, gevestigde aanduiding van die trappe van vergelyking in die AWS en sommige Afrikaanse woordeboeke is deur afleiding met *-er* of *meer* —, *-ste* of *mees -e*, byvoorbeeld *deurdringend*, *-er* of *meer* —, *-ste* of *mees -e*. Die aanduiding van die oortreffende trap met weglating van die lidwoord *die* is ten beste misleidend en ten slegste foutief. Dit vorm 'n integrale deel van die oortreffende trap in Afrikaans. Dit is jammer dat dié puik voorbeeld nie deurgedra is na bl. S15 vir die voorbeeld *goed*, *beter*, *beste* nie.
- (b) Voorbeeldsinne is oor die algemeen goed gekies en in die meeste gevalle laat dit die betekenis duidelik blyk. Voorbeelde is: **boord**¹ — Daar is 'n klomp appelbome in die **boord**. *There are a lot of apple trees in the orchard*; **redakteur** — Wie is die **redakteur** van daardie koerant? *Who is the editor of that newspaper?* Enkele voorbeeldsinne is egter niksseggend en kan met 'n bietjie aandag verbeter word, byvoorbeeld **skirt** (Engels) — My new **skirt** is very pretty. *My nuwe romp is pragtig* (wat van talle ander voor-

werpe gesê kan word). Iets soos: *My nuwe romp pas by my bloes* sou die betekenis beter laat blyk het. (Dié voorbeeld is van die Engelse kant aangehaal omdat die voorbeeldsinne nie aan die Afrikaanse kant ingesluit is nie — 'n verskynsel waarvoor hieronder meer gesê word.)

- (c) Uitdrukkings is tot die essensiële beperk, byvoorbeeld **van gunste en gawes leef, tussen hakies, kans vat, per ongeluk, net vel en been wees, stille vennoot** en dergelike meer.
- (d) Opmerkings aan die einde van sekere lemmas gee goeie leiding, byvoorbeeld by **meubels** 'n nota dat die enkelvoud **meubelstuk** is; by **pens** dat dié woord net in verband met diere gebruik word.

In die voorwerk onder **Woordeboekkenmerke/Dictionary features** (bl. v) word aangedui dat voorbeeldsinne by bykans elke trefwoord gegee word. Dit is inderdaad so, maar daar is hiate wat moeilik verklaar kan word. Byvoorbeeld: **blokfluit** het geen voorbeeldsinne nie, die Engelse **recorder** wel en in albei tale; dit geld ook **drom, dwarsfluit, ekologie, ekwivalent, inmeng, koalsuurgas, korrupteer, kraag, kous, marimba, masjien, nuuskierig, ontsteking, portuur, snoesig** en andere. Die teenoorgestelde geld ook: **monster** aan die Engelse kant het geen voorbeeldsinne nie, die Afrikaanse **monster** het wel en in albei tale.

Die opname van **metafoor/metaphor** sonder voorbeeldsinne het 'n geleentheid laat verbygaan om 'n moeilike taalkundige begrip vir leerders te verduidelik. Die oorbekende voorbeeld kon maar weer ingespan gewees het: *As 'n mens die kameel die 'skip van die woestyn' noem, is dit 'n metafoor; die kameel is nie 'n skip nie, maar soos 'n skip.* Dit geld ook die afwesigheid van voorbeeldsinne by **vergelyking/simile**.

Die morfologie van sekere lemmas verskil van die AWS, byvoorbeeld **fliëk** het as meervoud slegs **fliëke**, nie ook **fliëks** nie. By tweelettergrepige werkwoorde op *-eer* is daar ook inkonsekwensies: **hanteer (het gehanteer), loseer (het geloseer), regeer (het geregeer), waardeer (het waardeer** — waar die **ge-** skielik ontbreek), maar **probeer (het probeer of het geprobeer)**. Hierteenoor is werkwoorde met drie of meer lettergrepe wat op *-eer* eindig heeltemal korrek (ge)hanteer, byvoorbeeld **het gefotokopieer, geïdentifiseer, geïgnoreer, geïllustreer, gekwalifiseer**. (By **kampeer** is die verlede tyd aangedui as **het gekamp** — sonder die *-eer*.)

Die indruk wat hierdie publikasie maak, is dat dit soos ander Oxfordwoordeboeke netjies en nuttig is en sy prys werd is; die leksikografiese ongelikhede behoort egter in 'n volgende uitgawe reggestel te word.

Anton F. Prinsloo
Johannesburg
(prinsa@worldonline.co.za)

M.O. Mbatha (Editor). *Isichazamazwi SesiZulu*. 2006, vi + 1353 pp. ISBN-13: 978-1-919959-84-9. ISBN-10: 1-919959-84-x. Pietermaritzburg: New Dawn Publishers. Price: R450 (Hb.), R250 (Sb.).

1. Introduction: Comprehensiveness/Representativeness of the dictionary

This is a significant and commendable contribution to the development of Zulu lexicography in general, and Zulu monolingual dictionaries in particular. There could not have been a better time for the publication of such a work dealing with a language whose lexicon expands daily. As stated by the editor, this work was started at the University of KwaZulu (oNgoye) in 1997 and resulted in two volumes edited by Prof. A.C. Nkabinde, namely *Isichazamazwi I* and *Isichazamazwi II* which appeared in 1982 and 1985 respectively. According to the editor, *Isichazamazwi SesiZulu* is an extension of the previous two volumes. However, because of the many changes and additions made to these volumes, it was decided to integrate these into the present compilation, consisting of about 21 000 entries.

2. Expected users

As the editor rightly claims, this dictionary will undoubtedly be an invaluable tool to everybody, from general users, to learners at school, university students, researchers and anybody in need of the explanation of words from different subject fields.

The compilers of *Isichazamazwi SesiZulu* have in good measure succeeded in this objective, for they have included

- words commonly used by people in their everyday speech.
- loan-words which people use in their daily encounters with others in various businesses and trades, e.g. *uthela* 'bank teller', *umeselane* 'builder', etc.
- neologisms referring to new concepts not previously known to the Zulu speech community, e.g. *umkhukhu* 'shack', etc.
- idiomatic expressions, e.g. *ukuthathela phezulu* 'to act high-handedly'; *ukubhekwa yidlozi* 'to find luck'; etc.
- proverbs, e.g. *Umuthi ugotshwa usemanzi* 'a child is disciplined whilst he/she is still young', *Ubude abuphangwa* 'Rome was not built in a day', etc.
- Latin equivalents for Zulu names of animals and plants, e.g. *uxamu* *Monitor niloticus*, *iminya* *Halleria lucida*, etc.

3. Lemmatisation

Word categories to which entries belong have been indicated as follows:

- nouns: **bz** (abbreviation for *ibizo*)
- verbs: **sz** (abbreviation for *isenzo*)
- interjectives: **bbz** (abbreviation for *isibabazo*)
- ideophones: **szk** (abbreviation for *isenzukuthi*)

The stems of the above-mentioned word categories, given in bold letters and arranged in alphabetical order, have been used for entries. For nouns, an italicised noun prefix has been added to each stem. The adding of the noun prefix will enable the user to identify the prefix with which a certain stem could be used. If a stem occurs with various prefixes, separate entries are made that show the various prefixes occurring with the specific stem, e.g. *í(li)daka* 'dry cattle dung', *ísídaka* 'black soil', *ú(lú)daka* 'mud', *úkú*daka 'to become drunk' and *úm(u)daka* 'heavy, brown bracelet, bestowed as royal honour'. However, since only a singular noun prefix is given with each stem, this might be problematic, especially to users who are not mother-tongue speakers, because they might not know what the plural form of the specific word is.

Additional information has been given in the form of symbols used next to words for the user to distinguish the various word forms, for example:

- to indicate words which have a singular form only, the marker † has been used, e.g. *ísinyama†* 'bad luck', *ú(lu)sinsi†* 'hair growing low on the forehead'.
- to indicate words which have a plural form only, the marker ‡ has been added, e.g. *ámánconzuncônzu‡* 'pieces taken off gradually', *ízishaqane‡* 'a lot of paper money'.
- to indicate loan-words, the marker * has been used, e.g. *í(li)fasíkoti** 'apron', *úmesisi** 'a white married woman'.
- to indicate *hlonipha* words (i.e. words that are avoided by certain sections of the speech community), the marker x has been added, e.g. *úkúkhíjāx* 'to walk', *ínethax* 'rain'.
- to indicate neologisms, the marker ° has been added, e.g. *í(li)bhoklóló°* 'a brave, confident male', *úbúcwēbe°* 'jewellery'.

Tone has also been indicated by means of the symbol ´ to mark high tone, e.g. *úkúgíya* 'dance', *ísíwula* 'fool'; and the symbol ^ to mark a combination of high and low tones, e.g. *úbulâwu* 'love potion', *í(li)hlûle* 'clot'.

Words which contain the syllabic nasal *m*, are identified by the inclusion of the elided vowel *u* written in brackets, e.g. *úm(ú)lotha* 'ash', *úm(ú)fula* 'river'.

If words have different pronunciations, these pronunciations are given,

e.g. *ímbokodo*(e) 'grinding stone', pronounced as *ímbokodo* or as *ímbokode*, and *ú(lú)bongiyâne* (i)^x 'bee', pronounced as *ú(lú)bongiyâne* or as *ú(lú)bongiyâni*.

All the verbal extensions of a particular verb have been given in parenthesis, e.g. *úkúlobola* (sz) [-an-, -el-, -ek-, -is-, -w-]. However, variants of words which are the result of phonological processes such as the passive construction have been omitted, e.g. *úkúlobola* which changes to *úkúlotsholwa* in the passive.

Definitions of words immediately follow grammatical information such as word category and extensions. If a word has more than one meaning, the various meanings, polysemous or homonymous, are numbered by bold Arabic figures and listed one after the other, e.g.

úkúfunda (sz) [...] **1.** ukufaka iqatha noma umthamo emlonyeni 'to swallow chewed food'. **2.** ukuhamba isikole uphokophele ukuthola ulwazi 'to go to school in order to gain knowledge'. **3.** ukuthamunda noma ukubuka lokho okubhaliwe bese ukuqeketha ngomlomo noma ukuthamunde buthule 'to read'. **4.** ukuvuleka ingqondo 'to be alert'.

If a word is synonymous, its synonym or synonyms are given after each definition, e.g. *úkúdina* is treated as follows:

úkúdina (sz) [...] **1.** ukubangela ukucasuka 'to cause a feeling of disgust'; ukucasula 'to disgust'; ukunenga 'to disgust'; ukucunula 'to disgust'. **2.** ukuqeda amandla 'to exhaust'; ukufakela ukukhathala 'to make tired'; ukukhathaza 'to tire'.

Each of these synonyms are lemmatised as separate entries and do not refer back to *úkúdina*, e.g. *úkucasula* (sz) ukwenza umuntu ukuthi athukuthele 'to make someone angry', and *úkúnenga* (sz) ukwenza umuntu acasuke 'to disgust someone'.

Definitions are supplemented by sentences in italics, to show how the words are used in context, e.g. to illustrate the use of the first sense of *úkúdina* in context, the following sentence has been given: *Ayangidina amanga akho*. 'Your lies disgust me' and the use of the second sense is illustrated by the sentence: *Umsebenzi esabe siwenza wasidina sonke ...* 'The work that we did exhausted us all ...'

If a word can be used figuratively, the idioms or proverbs formed with the specific word are given. Anyone consulting the dictionary will easily identify idiomatic expressions marked by the abbreviation *ssh* (*isisho*), e.g. *ukudla amathambo ekhanda* (*ssh*) 'to think', *ukudla amabele* (*ssh*) 'to be alive', and proverbs indicated by the abbreviation *sg* (*isaga*), e.g. *Inja iyawaqeda amanzi ngolimi* (*sg*) 'perseverance leads to success' (*Inja*) *ithi ingahamba udle udaka* (*sg*) 'a hungry person can do anything.'

4. Conclusion

The compilers of this dictionary should be congratulated on this achievement in Zulu lexicography. However, it should be noted that technology has brought with it many terms, such as the computer-related terms *mouse*, *e-mail*, *internet*, etc., which have not been included in the dictionary. This should be borne in mind when a revised edition of the dictionary is produced. In view of the size and extent of the dictionary, it would be of great help to users if a CD-ROM could accompany the paper dictionary. This would be in line with modern trends.

Rose Masubelele
Department of African Languages
University of South Africa
Pretoria
Republic of South Africa
(masubmr@unisa.ac.za)

Anton F. Prinsloo. *Sleng: Woorde, uitdrukkings en hul herkoms.* 2006, vii + 256 pp. ISBN 978 186891 813 3 (1 86891 813 0). Kaapstad: Pearson Education South Africa (Pty) Ltd. Prys: R185.

As die jongste toevoeging tot Anton Prinsloo se bruikbare naslaanwerke oor Afrikaans wat terselfdertyd aangename leesstof is, vul *Sleng* 'n leemte in die boekstaving van die Afrikaanse omgangstaal. Soos met o.a. *Feite, flaters, fiksie: 'n Opgaaf van die ware, die onware en die onwaarskynlike* (2001) en *Spreekwoorde en waar hulle vandaan kom* (2004) is *Sleng* nie alleen op taalpraktisyns gemik nie, maar is dit ook baie toeganklik vir die gewone taalgebruiker. Die boek bied eerstens 'n omvattende versameling woorde en uitdrukkings wat tans as omgangstaal (en dus substandaard) gereken word, maar gee daarnaas ook die "korrekte" Afrikaanse ekwivalent en die Engelse vertaling daarvan. Interessante feite oor die herkoms van die slengwoorde en -uitdrukkings skep boonop die geleentheid vir 'n boeiende leeservaring.

Reeds in die 2002-uitgawe van die *Afrikaanse woordelys en spelreëls* (AWS) is erkenning verleen aan Omgangsafrikaans — die informele variëteit van Standaardafrikaans — met die insluiting en bereëling van 500 woorde wat moedertaalsprekers in gemoedelike omgang met mekaar gebruik en waarvan die meeste Engelse ontlennings is. Die ongeveer 4 000 inskrywings in *Sleng* is 'n waardevolle voortsetting en uitbreiding van dié AWS-lys. Woorde en uitdrukkings van Omgangsafrikaans wat in die jongste Afrikaanse woordeboeke (die *Verklarende handwoordeboek van die Afrikaanse taal*, vyfde uitgawe, en die *Afrikaans-Engels/Engels-Afrikaanse woordeboek* van Pharos) opgeneem is, en derhalwe by implikasie tot Standaardafrikaans gereken word, is nie in *Sleng* ingesluit nie.

Die bondige "Toeligting" bied waardevolle inligting oor o.a. die soorte inskrywings, inligtingstipes en tipografie wat dit vir die gebruiker maklik maak om die stof optimaal te ontsluit. Interessant is veral die vermelding van die vier soorte Afrikaanse slengwoorde en -uitdrukkings wat blyke van die aard van taalbeïnvloeding gee. Daar is bv. dié wat heeltemal in Afrikaans vertaal is (soos *blomkoolore* en *die moeder van alle*), dié wat gedeeltelik vertaal is (soos *inchip* en *op die bandwagon klim*), dié waarin die oorspronklike Engels net so in Afrikaans gebruik word (soos *mash* en *jack of all trades*) en uitspraaksleng (waar die Engelse in plaas van die Afrikaanse woordklem gebruik word), bv. "dé-fek" en "kón-tak" in plaas van "de-fék" en "kon-ták" onderskeidelik. In die teks self word ook voorbeelde gegee waarby nie net die Engelse woordklem nie, maar ook die Engelse klankvorm oorgeneem is, bv. "monjoement" in plaas van "monument". Verafrikaanse slengvorme, bv. *kondem*, *liekeriesj*, *paartie*, *rob-bies* en *sjaddap* word nie as aparte trefwoordkategorie in die "Toeligting" vermeld nie, maar is volop in die teks. Waar die verafrikaansing nie tot die skrifbeeld deurgedring het nie, word nuttige uitspraakleiding gegee, bv. by *paralytic* wat as "paralêतिक" uitgespreek word en *gravel* waarby die Engelse [æ]-klank met 'n Afrikaanse [ɛ]-klank weergegee word.

In die "Toeligting" word verder die verskillende inligtingstipes genoem wat binne 'n inskrywing kan voorkom. Naas die trefwoord verskyn die slengwoord of -uitdrukking, waar toepaslik, soos wat dit in Afrikaans gebruik word. Indien nodig, word die trefwoord van 'n gebruikseeraanduiding of -waarskuwing voorsien. Waar die trefwoord nie Engels is nie, word die Engelse ekwivalent gegee. Daarnaas volg die idiomatiese of Standaardafrikaanse ekwivalent. Indien bekend, volg 'n bespreking van die herkoms van die woord of uitdrukking. 'n Voorbeeldsin word soms ter illustrasie gebruik en waar nodig, kom verwysings na ander trefwoorde voor. Die gebruiker wat bv. die Engelse trefwoord *boot*² naslaan, kry die gebruikseeraanduiding "rekenaar", gevolg deur die Standaardafrikaanse ekwivalent "selflaai". Dan volg 'n verduideliking dat *boot* 'n inkorting is van *bootstrapping*, 'n vroeëre proses waarby 'n klein hoeveelheid data gekodeer en aangewend is om progressief komplekser kodes te laai, totdat die apparaat geskik was vir gebruik — aldus die herkoms van die Engelse woord. 'n Kruisverwysing tussen die trefwoorde *boot*² en *skoenveters*, bloot omdat *skoenveters* 'n slengvorm van *bootstraps* is, het nie veel waarde nie. In teenstelling hiermee is die kruisverwysing tussen *Pommie/Pommy* en *limey*, beide neerhalende benamings vir *Engelsman*, nuttig. Ook wat die herkoms van dié benamings betref, is daar 'n ooreenkoms: *Pommie* was oorspronklik 'n Australiese bynaam vir 'n Engelsman omdat die Engelse soldate granate "pomegranates" op die skepe geëet het vir die vitamine C-inhoud daarvan wat skeurbuik voorkom het, terwyl die Oos-Indiese Kompanjie om dieselfde rede aan Britse matrose suurlemoen- en lemmetjiesap gegee het, wat daartoe gelei het dat die matrose *lime-juicers* genoem is, wat later tot *limeys* verkort is.

Afgesien daarvan dat *Sleng* aan taalpraktisyns en gewone taalgebruikers leiding gee oor wat tans as substandaardvorme beskou word, is dit ook 'n skatkie van treffende neologismes in Afrikaans. Word die Afrikaanse nuutvorming teenoor die Engelse slengvorm gestel, staan 'n mens dikwels verstom oor die seggingskrag van Afrikaans en wonder jy onwillekeurig waarom omgangtaal-sprekers dit nodig vind om die taal met slengvorme te verarm. Vergelyk in dié verband die volgende: *hangparlement* "hung parlement" teenoor *skaakmatparlement*, *kasliberaal* "closet liberal" teenoor *doekvoetliberaal*, *tie-dye* teenoor *knoop-endoop*, *Indiaanse somer* "Indian summer" teenoor *opslagsomer*. In ander gevalle weer, soos by *nerd*, noem Prinsloo tereg dat hierdie Engelse woord maar in Afrikaans opgeneem behoort te word. Die lys alternatiewe Afrikaanse benamings waaronder *melkpensie*, *bleeksiel*, *skaam outjie*, *japie*, *gawie*, *doffel* en *dikkop* weerspieël immers die skynbare onvermoë van Afrikaanse taalkundiges, skrywers, woordeboekmakers e.a. om 'n geskikte vertaling te kry. Dieselfde geld ook die Engelse kapselsamestelling *brunch*, uit *breakfast* en *lunch*, wat ingeburger is in Omgangsafrikaans.

Hier en daar kom terloopse "regstellings" op Afrikaanse woordeboeke se hantering van slengvorme voor. By *lammie* word bv. daarop gewys dat Afrikaanse woordeboeke die uitdrukking *iemand 'n lammie gee* verklaar as inkorting afgelei van *lamboud*. Soos dit met die Engels "hit someone causing a lame feel-

ing" die geval is, is dit nie net 'n boud nie, maar ook 'n arm wat lam gemaak kan word.

Die byderwetse bladuitleg en lighartige illustrasies van Frans Groenewald pas by die onderhoudende styl van dié baie leesbare taalbron. Die tipografiese konvensies, soos in die "Toeligting" vermeld, dra verder by tot die gebruikersvriendelike formaat. Daar word tussen Afrikaanse en Engelse trefwoorde onderskei deurdat eersgenoemde in vet romeins (bv. **spietkop** en **strate bo iemand wees**) en laasgenoemde in vet kursief (bv. *repossess* en *the sky is the limit*) gegee word. By hibriede trefwoorde (bv. *liftklub* en *share en share alike*) word die Afrikaanse deel vet romeins en die Engelse deel vet kursief aangebied. Die Engelse weergawe word tussen ronde hakies geplaas, bv. (traffic officer), (be streets ahead of someone), (lift club) en (share and share alike). Die Afrikaanse (idiomatiese) ekwivalent word tussen vierkantige hakies geplaas (bv. [verkeersbeampte], [stukke beter wees as iemand, iemand ver oortref], [terugneem], [daar is geen perke nie], [saamryklub] en [gelykop verdeel]). Hier en daar, bv. by *pie in die sky*, word nagelaat om die Engelse weergawe "pie in the sky" te gee, terwyl tipografiese foute soms begaan word, bv. by die Engelse trefwoord **quickstep** wat vet romeins is, terwyl dit vet kursief moes wees, maar oor die algemeen is die tipografiese versorging keurig.

Aparte inskrywings onder 'n trefwoord word vet romeins genommer en waar daar 'n reeks uitdrukkings as aparte inskrywings voorkom, word die kernwoord in elke uitdrukking onderstreep en die uitdrukkings alfabeties volgens kernwoorde gerangskik, bv. blaas, forseer, gee, hê, ingooi, inhou, inkry, ken, mond, oorspeel, ou, pie en vry by **hand**. Hierdie hanteringswyse dra grootliks by tot die vindbaarheid van leksikale items.

Met die ⇒-teken word verwysings na ander trefwoorde aangedui, bv. *DJ* wat vir bewerking na *disk jockey* verwys word en *straight* wat na *royal 2* verwys word omdat hulle verwante pokerterme is. Hierdie tipografiese merker is meer gebruikersvriendelik as bv. 'n "Sien"-verwysing.

Die alfabetiese volgorde van die trefwoorde is woord-vir-woord. Meervoudsvorme van trefwoorde is deurgaans as aparte inskrywings opgeneem (bv. **vriend**, **vriende** en **lyn**, **lyne**) en verhewe syfers word gebruik om trefwoorde wat tot verskillende woordsoortelike kategorieë behoort (bv. *pick*¹ s.nw. en *pick*² ww.) van mekaar te onderskei. Hierdie hanteringswyses dra daartoe by om soektogte te vergemaklik.

'n Verdere pluspunt is die insluiting van 'n Engelse register agterin die boek. Hiermee word die stof ook vir Engelssprekendes toeganklik gemaak. Sou 'n Engelse gebruiker nie die Engelse woord waarna hy soek op sy alfabetiese plek in die hoofteks vind nie, kan hy die register raadpleeg om te bepaal onder watter Afrikaanse trefwoord hy die betrokke woord kan terugvind. As die gebruiker bv. *play* naslaan en dit nie op sy alfabetiese plek in die hoofteks kry nie, word hy in die register na die trefwoorde **speel**, **rondspeel** en **uitspeel** verwys waarin daar van *play* melding gemaak word.

As eerste naslaanbron oor Omgangsafrikaans verteenwoordig *Sleng* baanbrekerswerk en as sodanig verdien dit 'n plek op die rak van elke taalpraktisyn, akademikus, opvoeder en leerder wat gesteld is op "korrekte" Afrikaans. Maar ook die gewone taalgebruiker kan hom verkneukel in die aardigheid van Afrikaans. 'n Mens kan maar net hoop dat die boek uiteindelik bygewerk sal word om meer trefwoorde in te sluit, want die *ins en outs* (*besonderhede of fyn puntjies*) van Afrikaans is nog lank nie volledig beskryf nie!

Bronne

- Du Plessis, M. (Hoofredakteur/Chief Editor).** 2005. *Pharos 1, Afrikaans–Engels/Engels–Afrikaans Woordeboek/Dictionary*. Kaapstad/Cape Town: Pharos Woordeboeke/Pharos Dictionaries.
- Odendal, F.F. en R.H. Gouws (Redakteurs).** 2005. *HAT. Verklarende handwoordeboek van die Afrikaanse taal*. Vyfde uitgawe. Kaapstad: Pearson Education South Africa.
- Prinsloo, Anton F.** 2001. *Feite, flaters, fiksie: 'n Opgaaf van die ware, die onware en die onwaarskynlike*. Kaapstad/Pretoria/Johannesburg: Human en Rousseau.
- Prinsloo, Anton F.** 2004. *Spreekwoorde en waar hulle vandaan kom*. Kaapstad: Pharos.
- Taalkommissie van die Suid-Afrikaanse Akademie.** 2002. *Afrikaanse woordelys en spelreëls*. Kaapstad: Pharos.

A.E. Cloete
Buro van die Woordeboek van die Afrikaanse Taal
Stellenbosch
Republiek van Suid-Afrika
(aec@sun.ac.za)

Włodzimierz Sobkowiak. *Phonetics of EFL Dictionary Definitions*. 2006, 249 pp. ISBN 83-7177-450-8. Anglistyka. Poznań: Wydawnictwo Poznańskie. Price: 38 zł.

I dream of dictionaries which would redress the anti-phonetic bias of current lexicography.

— Sobkowiak 2006: 11

[A] phonetically uncontrolled defining vocabulary could breed a cancerous monster of a text.

— Sobkowiak 2006: 68

Just as book reviews are written for a variety of reasons, the reviews themselves serve various purposes. Chief among the latter is probably one's wish to know whether or not one should actually *read* the book under review. Let me therefore state upfront that, yes, this work should be on every (meta)lexicographer's bookshelf, where it should only be allowed to gather dust once it has been carefully studied and its contents seriously pondered. When Sobkowiak himself claimed, on p. 14 of the work under review, that "every lexicographer should read this book", I was sceptical, but after having read through the entire text twice, I have to conclude that he is indeed correct.

The reason for this is not so much the fact that this work contains valuable phonetic data (it does, yet that is a given), but rather the fact that core lexicographic principles and methods are presented, summarised and reanalysed in the process. Sobkowiak himself is a man of wide reading, and it shows. Very few aspects needed for his exposition escaped him. What he does need, however, is a PR-officer. As it stands, very few (meta)lexicographers will take to a monograph with the exotic title "*Phonetics of EFL Dictionary Definitions*". Had that title just been "EFL Dictionary Definitions", or even "Dictionary Definitions for Learners of English — How They Can Be Improved", the book would have been an instant bestseller. By doing so one is not tricking the reader, rather, one makes sure that that reader is not put off before even picking up the book. (Admittedly, the title of the current book is already better than that of Sobkowiak's 1999 monograph: "Pronunciation in EFL Machine-Readable Dictionaries", where something (again equally relevant) like "Electronic Dictionaries: The State of the Art & The Shape of Things to Come" would again have meant instant success.)

Of course, with close to three decades of research into inter alia English phonetics and phonology as well as English lexicography — making him one of the world's sole phonolexicographers — I understand why it is hard for Sobkowiak to give in to market pressures. But when one has something to say (as is the case for Sobkowiak), one must also attempt to get the message across.

Sobkowiak's message in this book is a fascinating one as, despite the current hype and excitement as a result of ever better concepts and tools put forward for dictionary making (with corpus-driven compilation a *sine qua non*), important concepts and tools *remain fully untapped*. In this book that concept is the notion that definitions (in learners' dictionaries) need to be controlled pho-

netically, so that these definitions can be processed more easily by the dictionary users. The suggested tool is the so-called Phonetic Difficulty Index (henceforth PDI). With lots of electronic (corpus and dictionary) data at one's disposal, PDIs can be studied on the word level (for example with regard to Defining Vocabularies (DVs)), or on the level of entire definitions (as well as on any other microstructural level).

The book is divided into three parts: a front section consisting of an Abstract and an Introduction, a middle section containing eight Chapters in which the main argument is logically developed, and a back section including a Bibliography, Table and Figure indexes, and seven Appendices. Opening a monograph with a page-long abstract is somewhat strange, especially when the first paragraph already mentions the end conclusion: "[D]efinition text can, and should, be to some extent phonetically controlled to avoid lexical maxima of high phonetic difficulty, ensure easy (subvocal) reading, hence — foster understanding and incidental vocabulary learning by learners" (p. 7). The Introduction goes on to lament the fact that "nothing much has changed in lexicographic phonetics since the advent of computers" and this, on top, "in the context of almost complete oblivion of matters phonetic in lexicography" (p. 10). The target of this book is defined as "mainly (meta)lexicographers and EFL teachers (and learners)" (p. 13).

Chapter 1 is mostly a presentation of relevant existing studies in EFL vocabulary learning on the one hand, and an enumeration of why phonetic difficulty should be addressed by dictionary makers at all on the other. In Chapter 2 the creation of a 'reference lexicon' (i.e. the lexicon, derived from OALD3, against which all other data is compared throughout the book) is described. In Chapter 3 the PDI itself is introduced, which turns out to be a tool that captures pronouncing difficulties by Polish learners of English (who end up speaking 'Polglish'). A total of 57 such difficulties are listed, and each is assigned a value of 1. The validation of the PDI tool as summarised in Figure 2 (p. 51) is especially appealing.

Chapter 4 proceeds with a detailed discussion of the phonetic aspects of the DVs of the so-called Big Five monolingual EFL dictionaries of English, viz. OALD, LDOCE, COBUILD, CIDE and MEDAL. Overall, DVs are shown to be "phonetically easier than general vocabulary" (p. 66). In Chapter 5, the (general) phonetic aspects of EFL dictionary definitions are then discussed. Unfortunately, quite some material that had already been presented in the Introduction is repeated here, which indicates that the whole of the Introduction had better be integrated into the middle section of the work.

Armed with all the knowledge and data from the previous Chapters, Chapter 6 then offers a detailed case study of the phonetics of MEDAL definitions. A wide variety of permutations of data drawn from the reference lexicon, the BNC, other Big Five dictionaries and MEDAL is effected, with often highly interesting (though predictable) results. In Chapter 7 several dozen pedagogical applications of phonetically treated definitions are then offered, while Chapter 8 concludes the study with a recap of the main outcomes.

One of these main outcomes of the study may be summarised with the following question: "If *thorough* is among the phonetically hardest lexical items for Polish EFL learners (PDI=6), for example, why not use a substitute (*complete* has PDI=1) in definitions [...]?" (p. 151). Likewise, on p. 90 it is suggested to use, wherever possible, *if* instead of *whether* in definitions, or on p. 102, to use *substance which* rather than *substance that*, "a phrase which would make the user's life easier". Given there are differences between *which* and *that*, and that, in general, there are no real synonyms, I doubt many definition writers would be all too happy with such changes though.

As can be seen, I have refrained from showing any phonetics, formulas or otherwise incomprehensible codes in this review. This was on purpose, as I wanted to make sure that Sobkowiak's core message would be conveyed. In this respect, Sobkowiak himself is not doing the reader a great favour in his book, however. One simply has to know quite a bit about lexicography already, as well as of phonetics and even statistics, in order to smoothly transit from one section to the next in his book. Especially taxing is the (at times haphazard) use of abbreviations for all dictionaries. They are not always standard, change throughout the text, and have not been brought together anywhere either. Those who do not know the dictionaries themselves and/or their abbreviations will have to read through the entire Bibliography in order to find their meaning. Similarly, the first time the reader encounters a phonetic script in the book (on p. 21), it is not the usual IPA, but an ASCII rendition. From there onwards, the ASCII renditions only become more and more complex. A table mapping the ASCII versions with the more usual IPA symbols would have been welcome in the back section. Even more obscure are the various codes used to represent parts of speech and then later PDI codes. Again, at least a single table in the back section should have brought the codes and their meanings together. Nor would it harm to expand, in say a glossary in the back section, corpus and statistical abbreviations such as LOB, BNC, KWIC or MI.

On the whole, these are but a few aspects of the general carelessness that permeates the entire text. Not surprisingly, the monograph brings together various results that have been published over the years, in different formats, in a range of journals, each with its own style. Best would probably have been to rewrite all the material, as the frequent self-quotes and self-references break the flow. In addition, while most of the text uses 'I/my', there are accidental remnants of 'the present author' (cf. e.g. p. 70). Some citations have also been repeated more than once (cf. e.g. p. 79 = p. 103). The text could also have done with an extra proofread, as especially the definite article is (and occasionally the indefinite articles are) missing. Several dozen spelling errors, typos and omissions may also be found, as the following sample shows:

- p. 59: tile → title
- p. 66: CV → DV (twice, plus various other errors on this page)
- p. 82: page 47 → page 65

-
- p. 88: A number of interesting observation → A number of interesting observations
 - p. 91: URL, extra white space should be deleted
 - p. 94: worth while → worthwhile
 - p. 123: was presented above in , it was used in above, the list in below
 - p. 129: was used once again collect → was used once again to collect
 - p. 153: The full potential ... can only be released ... → The full potential ... can only be realised ...

Furthermore, the meaning of "[*]" in the Bibliography is only explained on p. 78; it could and should at least also have been mentioned at the start of the Bibliography itself. Given the Appendices are linked to very specific pages in the middle section, mentioning the respective pages could similarly have eased readability, thus:

- Appendix 1: p. 34
- Appendix 2: p. 49
- Appendix 3: p. 60
- Appendix 4: p. 61
- Appendix 5: p. 99 (p. 96)
- Appendix 6: p. 131
- Appendix 7: p. 135 (p. 132, 140, 143)

One of the dangers of recycling one's own earlier texts is that one may also forget to explain some core terms. This is the case with the frequently used 'phonolapsological' — which is not clarified anywhere, even though it is used throughout the monograph. Googling for more information does not help, as the nine hits all refer to titles of works by Sobkowiak himself, cf. below: "(9, all WS)", where also other Sobkowsiesque words used in the book are listed:

- non-bi-uniqueness / non-biuniqueness (2)
- phoncordance (2, both WS)
- phono-didactic (0)
- phonolapsological (9, all WS)
- PhonoLapsological Affinity Classes (0)
- PhonoLapsological Identity Class (1, WS)
- "phonolexical metric" (0)

In a new edition of the book, all of the above infelicities could easily be attended to. Three more issues also seem worthwhile to address. Firstly, the discussion of type/token ratios on p. 95 may be dropped, as they are mostly meaningless as long as one does not work with *standardised* type/token ratios. Secondly, I would have appreciated a more critical stance on the claimed DV sizes, versus the real number of words used, as revealed through this study (cf. e.g. p. 62). Thirdly and lastly, how portable are the results from this study to other languages, and thus away from Polish? As it stands, this is now an open question (on p. 183), which seems unsatisfactory. Once one has an answer to this, the results will truly have the potential of being of 'universal' use.

Abbreviations used (numbers that follow dictionary abbreviations are edition numbers)

ASCII	American Standard Code for Information Interchange
BNC	British National Corpus
CIDE	Cambridge International Dictionary of English
COBUILD	Collins Birmingham University International Language Database
DV	Defining Vocabulary
EFL	English as a Foreign Language
IPA	International Phonetic Alphabet
KWIC	Keyword in Context
LDOCE	Longman Dictionary of Contemporary English
LOB	London/Oslo/Bergen Corpus
MEDAL	Macmillan English Dictionary for Advanced Learners
MI	Mutual Information
OALD	Oxford Advanced Learners' Dictionary
PDI	Phonetic Difficulty Index

Reference

Włodzimierz Sobkowiak. 1999. *Pronunciation in EFL Machine-Readable Dictionaries*. Poznań: Motivex.

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
TshwaneDJe HLT
Pretoria, Republic of South Africa
(gillesmaurice.deschryver@UGent.be)

Birgit Steinbügl. *Deutsch-englische Kollokationen: Erfassung in zweisprachigen Wörterbüchern und Grenzen der korpusbasierten Analyse.* 2005, VI + 189 pp. ISBN 3-484-39126-X. ISSN 0175-9264. Lexicographica. Series Maior 126. Tübingen: Max Niemeyer. Price: €98.

This study investigates the role of collocations in dictionary use, and the extent to which users' needs are taken into account in the process of dictionary writing. Steinbügl decided to concentrate on bilingual dictionaries, because this type of dictionary is relatively less well explored in metalexicographical literature. German-English examples are analysed and evaluated. Instead of selecting examples randomly, she uses a comparative corpus of 200 collocations she put together herself in accordance with scientific reasons explained in detail. She questions the selection of collocations from existing corpora for her purposes, because these corpora are based on competing collocational theories. In order to come to meaningful conclusions, she prefers to delineate her own research approach, also however investigating the structures of bilingual dictionaries and dictionary articles, as well as situations of dictionary use.

In the first chapter of her book, Steinbügl gives a theoretical exposition by discussing different views on the notion of "collocation" (2005: 3). She firstly explains Hausmann and others' approach, which she describes as a "lexicographically-didactically oriented theory" (2005: 12). Hausmann distinguishes between non-fixed and fixed combinations of words and figures of speech. A fixed combination is encountered when its meaning cannot be explained by simply considering the semantic values of the individual components; this means that fixed expressions are to be interpreted as one sign. For example, the expression *chambre forte* is not automatically understandable for non-native speakers of French, because they cannot deduce the meaning from the individual components in the expression. Knowing the meaning of the individual words does not help them to interpret the fixed expression, which applies to the receptive as well as the productive use of a foreign language. For dictionary users, it is sometimes difficult to locate the explanations of fixed expressions, especially when dictionaries differ in their methods of placing these expressions.

Hausmann (1984: 399) describes three types of non-fixed combinations. Firstly, there are free combinations or co-creations, which can be produced at any time, and in which each individual component has an independent meaning. Secondly, he identifies collocations or so-called "affine combinations". He regards "affinity" as the "tendency of two words to appear in combination" (1984: 398). Collocations are not creatively formed; rather, they are "pieces" memorised by the speaker because they so often occur in combination. As the combinatory possibilities of the collocation partners are restricted, the use of the words are considerably reduced. It is therefore not possible to adequately explain the meanings of these combinations by semantic means. Thirdly, there are so-called "contra-creations" which occur only rarely, e.g. when novelists

deliberately break semantic and/or grammatical rules in order to combine words normally not combinable, usually for metaphorical effect.

Steinbügl draws strongly on Hausmann's view, especially using his second category as her point of departure. She emphasises (2005: 6, note 6) that the components of collocations need not be juxtaposed together in texts; Hausmann (1985: 124ff) accordingly uses the expression "collocation span" to indicate that words appearing relatively far away from each other in a sentence could still form a collocation. He also differentiates between "basis" and "collocator". Steinbügl explains this difference with *guilty conscience* as example. The noun *conscience* can be understood as the basis, and the adjective *guilty* as the collocator which completes the collocation. The basis which is semantically autonomous, can be defined without a context; the collocator can only be defined in terms of the basis.

From another viewpoint, the so-called "statistically-oriented approach", propagated by linguists such as John Sinclair and Göran Kjellmer, the presence of collocations is indicated when two linguistic entities appear together in the same text. When this simultaneous occurrence is found more often than expected, it is called a "significant collocation". This statistical approach is closely linked to computer-assisted corpus linguistics. In her study, Steinbügl also analyses corpus-based collocations, even though she believes that frequency alone cannot be taken as criterion for the identification of collocations. She proceeds from the assumption that the intuition of the linguist or lexicographer cannot be replaced by machine-based investigations, even though data obtained by computers can be more reliable than the intuition of a single mother-tongue speaker.

In British contextualism, Firth claims that the situational context in which an utterance is made, is of crucial importance to the meaning of language. Apart from this extra-linguistic dimension, meaning also has a pure linguistic level, on which collocations play a decisive role. He uses the notions of "habituality" with which words occur together, and the "mutual expectancy" of words to appear together. Differing from Hausmann's view that a hierarchical relation exists between basis and collocator, he believes that the meaning of both partners in the combination is equally influenced by the collocation.

The "text-linguistic" notion of collocation represented by scholars such as Halliday and Hasan uses the concept "collocational cohesion", which can be exemplified by *hair, comb, curl* and *wave*, which, as extra-linguistic entities, would form a combination and therefore a collocational cohesion. According to Steinbügl (2005: 15), however, the fact that no distinction is made between linguistic and extra-linguistic relations is problematic.

One of the problems facing lexicographers of bilingual dictionaries is that different languages have different ways of forming collocations. This also makes it extremely difficult for foreign language learners. Another aim of Steinbügl in this study is not only to look at pairs of two words, but also to investigate idiomatic and fixed expressions.

After the initial exposition of her theoretical considerations, Steinbügl discusses the theoretical foundations of bilingual dictionaries in the second chapter. She gives brief overviews of structural aspects such as the macrostructures of bilingual dictionaries and the microstructural components in bilingual dictionary articles. She also considers problems which may arise in the writing and use of dictionaries. She draws examples for her analyses from the following bilingual dictionaries: Duden Oxford bilingual dictionary (DOG-E2), Langenscheidts' bilingual dictionary with German and English (LHW-E 2000), Langenscheidts' bilingual dictionary with German and French (LHW-F 2000), Pons Collins dictionary with German and English (PCGW-E4) and Pons bilingual dictionary with German and French (PGW-F1). Later in her study, she then uses features of these examples for quantitative analysis.

Steinbügl analyses the niche-alphabetical macrostructures from some of these dictionaries. In the LHW-E 2000, the headwords within the niche-alphabetical grouping are provided with a grey background to help users find these within the dictionary article. Other dictionaries, such as the PCGW-E4, use one component of the headword (e.g. *Vorsichts-*) to indicate the beginning of a particular niche. Even though nest-alphabetical grouping may help in the learning function for users, it may be more difficult and time-consuming for users to find the looked-for headwords. Consequently, niche-alphabetical arrangements seem to be preferable.

With regard to the microstructures of bilingual dictionary articles, Steinbügl acknowledges the well-known types, namely, integrated, non-integrated and partially integrated microstructures. The different meanings of a headword are usually introduced typographically with alphabet letters or Arabic numerals preceding each separate meaning. In an integrated microstructure, collocations and other multi-lexical units are placed within the polysemic structure, which means that they are presented as part of the information given under each alphabet letter or Arabic numeral. In the case of non-integrated microstructures, multi-lexical units are expanded, which means that they are dealt with in a separate part at the end of the article. Partially integrated microstructures occur when multi-lexical units are sometimes treated in the polysemic structure, and sometimes in an annex designed especially for phraseologisms.

The lexicographical units which have to be dealt with in bilingual dictionary articles differ from those in monolingual dictionaries. In the bilingual dictionary article, each source-language address has to be assigned a target-language item. What makes it difficult, however, is the question whether the dictionary is being compiled for passive foreign language learners, i.e. with a view to understanding foreign language texts, or for active speakers, i.e. with a view to produce texts in the foreign language. Certain information could be superfluous for native speakers, but essential for foreign speakers. Other problems arise with words having multiple meanings, where the presentation of certain data could lead to repetitions. Too much information could also result in users experiencing confusion and despair.

Another problem that she identifies is the way many bilingual dictionaries present collocations under so-called "examples". According to Steinbügl (2005: 25), collocations or syntagms which are placed as examples do not illustrate the use of the headword in question, but rather give the translation of the collocation. An example is the German *mit den Zähnen knirschen*, which can be found under *knirschen* in bilingual dictionaries. This example, being a collocation translated as *grind one's teeth*, does not illustrate the use of *knirschen* (which means "crunch", used in connection with sand, gravel or snow), but in fact represents a figurative meaning of it. Steinbügl (2005: 26) pleads for a micro-structure in active bilingual dictionaries which is rich in collocations, because these multi-lexical units are very important for foreign text production.

She also looks at the arrangement of longer articles with regard to the ease with which users can obtain an overview. Firstly she finds that many bilingual dictionaries base the formation of the articles on semantic criteria, even though the approach is rather arbitrary. When the article for the adjective *ruhig* is observed, for example, it is found to be divided into five semantic groups in the PCGW-E4 and into four in the DOG-E2. Nevertheless, both these articles are fairly clear in their presentation of the information, in contradistinction to the LHW-E 2000 in which the semantic groups are typographically not presented in a user-friendly manner.

Secondly, other dictionaries follow a so-called "categorical-alphabetical principle", such as the Dutch publisher Van Dale's bilingual dictionaries. Here, the noun partner in the collocations is matched with the other partner (which can belong to nine different parts of speech, always numbered from 1 to 9, and always presented in the same order). Steinbügl (2005: 28) thinks that this type of presentation may be too complicated for inexperienced foreign language users, who might not be in a position to classify words according to their parts of speech. She quotes Van der Meer (1998: 222), who states that this type of presentation is a "*lexicographer's* solution to a *lexicographer's* problem and not one tailored to the *user's* problem and the *user's* abilities". Thirdly, in dictionaries without any codes, text blocks or other formal ways of helping users typographically, their searches may be much more difficult and less successful.

In a distinction between monodirectional and bidirectional dictionaries, the latter could have disadvantages because it will always be a compromise to speakers of both language groups. Steinbügl (2005: 35), however, maintains that most bilingual dictionaries are bidirectional. Referring to Hartmann, who states that information in bilingual dictionaries should not be symmetrical, but rather aimed at one particular language group, she claims that the direction of a dictionary can be inferred from the language of the metatexts (e.g. the preface of the DOG-E2 is exclusively in German). She criticises this dictionary by saying that in the articles themselves, the choice of the language for the metatexts is not based on scientific arguments, and when the German verb *stibitzen*, for example, is consulted, the stylistic markers such as "colloquial" given for the target-language equivalents are useful for German, but irrelevant to English

users. She cites many similar examples which result from this type of compromise.

With regard to the functions of bilingual dictionaries, it is important that dictionaries for active use (i.e. the production of texts) should not only give strings of target-language equivalents. Foreign language users are unable to select from these lists the equivalent appropriate for the context. Steinbügl also refers to other scholars such as Mugdan who propagated a series of six bilingual dictionaries for each language set, and who believed that for the production of a foreign language text, users should preferably use a good monolingual dictionary in the foreign language rather than consulting a bilingual dictionary.

After this general discussion of bilingual dictionaries and the various points of view held by scholars, Steinbügl considers collocations as they are found in bilingual dictionaries. She asks the valid question regarding the place where collocations should be listed: under the basis or under the collocator, or under both? The answer will always be influenced by the goal of the dictionary. The best practice is, of course, to list such collocations and their translation equivalents in four places, namely, under the basis as well as the collocator of both language pairs. This is, however, not possible for reasons of space, which means that the lexicographer should try to determine the most likely places where users will look for a particular collocation. Cross-references can then be used, although lexicographers should avoid a too expanded cross-reference system, which may cause confusion and user-unfriendliness.

Sometimes lexicographers use "reduced collocation items". For example, in the case of *Verwirrung stiften*, the PCGW-E4 lists the source-language collocation partner *Verwirrung* under *stiften*, but only presents *cause* as translation equivalent and not the full collocation *cause confusion*. According to Steinbügl (2005: 51), this is very user-unfriendly.

In the third chapter of her book, Steinbügl explains the criteria for her selection of dictionaries to be analysed and the compilation of the corpus to be used in her comparative investigation. Her comparative corpus will consist of collocations of which the translations into English cause problems for German speakers. Discussing the possibilities of statistically investigating collocations, she expresses her concern that the lexicographer's intuition should not be ignored totally. It is possible that machine-driven searches for and analyses of collocations could include word combinations which co-occur accidentally, and by only using such searches, useless results could be obtained. Furthermore, computers cannot (yet) distinguish between the different meanings of linguistic signs. Therefore, she recommends that in the search for collocations, the linguist's intuition should in the first place be trusted, and corpora only used in the second place to confirm this intuition.

Several text corpora are briefly introduced and discussed by means of examples, such as the ones from the Institut für Deutsche Sprache (IDS), the COBUILD Bank of English and the British National Corpus (BNC). A comparison between English and German corpora highlights the different compositions of such databases.

For her comparative corpus, Steinbügl also decided to use monolingual dictionaries for the identification of collocations. Her preliminary corpus consists of 200 collocations, listed on pages 85-89 of her book. Interesting examples of non-equivalence are cited, e.g. *einen Brief einwerfen* or *Staub wischen* (both collocations in German) as opposed to *to post* or *to dust* (no collocations in English). Even the frequency of collocations, when tested against the corpora, sometimes differ rather drastically. For example: The combination *ein Geschenk auspacken* occurs over a hundred times in the IDS, but *to unwrap a present* is found only once in the BNC. When the frequency of *to open a present* is investigated, it appears 31 times, which leads to the conclusion that *to open a present* is a better equivalent for the German collocation. Steinbügl lists many of this type of examples, and discusses some of the problems they cause. Therefore, after taking all kinds of problems into consideration, she sets up a final list of collocations, on pages 103-107, based on scientific argumentation.

The fourth chapter of Steinbügl's book contains her quantitative analysis of the collocational practice in German-English dictionaries. For this purpose, she compares the LGW-DE 1982, the DOG-E2 and the PGW-E1, using graphs to illustrate many results. She looks at the types of information given, the typographical designs, and other features in the presentation of collocations. One of her findings indicates that, in accordance with her expectations, most of the collocations can be found, not under the base, but under the collocator. She discovered that the DOG-E2, which is much more comprehensive than the other dictionaries investigated, gives much more useful information on collocations. This was, of course, also to be expected. The DOG-E2's collocational entries contain many so-called "reduced" collocations, but in its German-English part, the bases are given in English, which is beneficial for German speakers when producing texts. This differs from the LGW-DE 1982, the PGW-E1 and the PCGW-E4, where the German base is given in English under the German collocator, which makes it easier for users to select the right equivalent, but forces them to translate the base into English themselves. When the dictionaries are considered from the point of view of the frequency by which they place collocations under the base, the LGW-DE 1982 ranks the highest and the DOG-E2 the lowest.

In the final chapter, Steinbügl states, however, that her findings are disappointing, because the investigated reference works only treated a relatively small part of the collocations in her comparative corpus comprehensively, or even prominently in bold print. From this, she concludes that there might be a need for special collocational dictionaries.

Steinbügl's book can be very useful for investigating other language pairs from bilingual dictionaries to see what kind of results can be obtained when her approach is used as model. Also, compilers of bilingual dictionaries will do well to note Steinbügl's concern with regard to collocations, and their relatively problematic treatment. Her argumentation, based on linguistic insights and broad experience from her analysis of so many collocations from so many dictionaries, should be taken into cognizance when compiling bilingual dictionar-

ies with other language pairs. The book is written in a German which is not too complicated to be understood by readers with a basic knowledge of German; the examples are clear, and the many graphic illustrations are very helpful.

References

- Hausmann, F.J.** 1984. Wortschatzlernen ist Kollokationslernen. *Praxis des neusprachlichen Unterrichts* 31: 395-406.
- Hausmann, F.J.** 1985. Kollokationen im deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels. Bergenholtz, Henning and Joachim Mugdan (Eds.). 1985. *Lexikographie und Grammatik: Akten des Essener Kolloquiums zur Grammatik im Wörterbuch* 28–30.6.1984: 118-129. Lexicographica. Series Maior 3. Tübingen: Max Niemeyer.
- Van der Meer, G.** 1998. How to Find the Correct Target Language Equivalent. Zettersten, A., V. Hjørnager and J.E. Mogensen (Eds.). 1998. *Symposium on Lexicography VII. Proceedings of the Eighth International Symposium on Lexicography May 2–4, 1996, at the University of Copenhagen*: 25-40. Lexicographica. Series Major 90. Tübingen: Niemeyer.

Dictionaries

- DOG-E2 = Duden Editors and Oxford University Press (Eds.). 1999. *Duden Oxford. Großwörterbuch Englisch: Deutsch–Englisch/Englisch–Deutsch*. Second Edition. Mannheim: Duden.
- LGW-DE 1982 = Messinger, H. and Langenscheidt Editors (Eds.). 1982. *Langenscheidts Großwörterbuch Englisch. Der kleine Muret-Sanders. Part II: Deutsch–Englisch*. Berlin/Munich: Langenscheidt.
- LHW-E 2000 = *Langenscheidts Handwörterbuch Französisch: Französisch–Deutsch/Deutsch–Französisch*. 2000. Berlin/Munich: Langenscheidt.
- LHW-F 2000 = *Langenscheidts Handwörterbuch Englisch: Englisch–Deutsch/Deutsch–Englisch*. 2000. Berlin/Munich: Langenscheidt.
- PCGW-E4 = Terrell, P., V. Schnorr, W.V.A. Morris and Roland Breitsprecher (Eds.). 1999. *Pons Collins Großwörterbuch für Experten und Universität: Deutsch–Englisch/Englisch–Deutsch*. Fourth Edition. Stuttgart: Klett.
- PGW-E1 = Cop, M. et al. (Eds.). 2002. *Pons Großwörterbuch für Experten und Universität: Englisch–Deutsch/Deutsch–Englisch*. Completely New Edition. Stuttgart: Klett.
- PGW-F1 = *Pons Großwörterbuch: Französisch–Deutsch/Deutsch–Französisch*. 1996. Completely New Edition. Stuttgart: Klett.

Maria Smit
Department of Music
University of Stellenbosch
Stellenbosch
Republic of South Africa
(msmit@sun.ac.za)

H.C. Viljoen (Redakteur/Editor). In medewerking met/In collaboration with N.F. du Plooy en/and S. Murray. *Kuberwoordeboek Afrikaans–Engels, Engels–Afrikaans/Cyber Dictionary Afrikaans–English, English–Afrikaans.* Onder beskerming van/Under the auspices of Vriende van Afrikaans. 2006, 518 pp. ISBN-13: 978-1-86919-133-7. ISBN-10: 1-86919-133-1. Pretoria: Protea Boekhuis. Prys/Price: R250.

Kuberwoordeboek is 'n opgegradeerde weergawe van die *Tweetalige Rekenaarwoordeboek* van N. Coetzee, N.F. du Plooy en A.D. de V. Cluver wat in 1985, voor die vestiging van die internet, verskyn het. As 'n mens *Kuberwoordeboek* die eerste keer sien, is jy verbaas dat 'n publikasie wat so ouderwets lyk, soveel hedendaagse terme kan bevat. Die ouderwetsheid is te wyte aan die Windows 95-rekenaarblokkie op die voorblad. Indien 'n moderner Windows XP-blokkie eerder gebruik is, sou die woordeboek sekerlik meer byderwets gelyk het.

Alhoewel die woordelys baie omvattend is, en elke denkbare rekenaarterm insluit van *surf* tot *spyware*, klink baie van die vertaalekwivalente wat verskaf word, vreemd in Afrikaans. Neem byvoorbeeld die Engelse term *emoticon* (enigene van daardie oulike gesiggies wat 'n mens deesdae in e-posse kan invoeg): Die Afrikaanse ekwivalent wat verskaf word, is *emotikon*. Tog wonder 'n mens hoeveel mense regtig van *emotikons* praat. Gebruik hulle nie veel eerder *gesiggies* in alledaagse taal nie?

Die selfde vrae kan gestel word rondom die bewerking van die werkwoord *google*, want ten spyte van die regstryd wat tans aan die gang is betreffende *google*, is dit die term wat die algemeenste gebruik word in Afrikaans. Tog word *google* glad nie eens as Afrikaanse vertaalekwivalent gelys nie, maar word slegs *soekswarf* gegee.

Alhoewel dit wel waar is dat leksikograwe nuwe terme mag skep wat hulle meen inslag sal vind, is dit belangrik dat die terme wat die algemeenste gebruik word, ook in ag geneem sal word. 'n Wesenlike vraag wat dus ontstaan wanneer 'n mens na hierdie nuutskeppings kyk, is wat hulle rakleef tyd sal wees. Gaan rekenargebruikers oor 'n jaar of wat boodskappe op hulle *werfjoernale* of op hulle *blogs* plaas? En gaan enigiemand weet waarvan 'n mens praat as jy verwys na 'n *klikkloot* (*click potato*)?

Hier is enkele ander voorbeelde van goeie of slegte vertalings wat in *Kuberwoordeboek* voorkom: 'n *bit map* is 'n *bispatroon*, 'n *password* 'n *wagwoord*, 'n *hacker* 'n *kapper*, terwyl 'n *cracker* 'n *kuberkraker* is. *Cyber bugs* (byvoorbeeld virusse) word vertaal as *kubergoggas* en 'n *flash disk* staan bekend as 'n *flitsskyf*.

Dit is jammer dat *Kuberwoordeboek* nie meer omvattende leksikografiese bewerkings verskaf nie. Dit blyk 'n woordeboek te wees wat daarop gerig is om vertaalekwivalente, en nie veel meer nie, te gee. As 'n mens byvoorbeeld kyk na die inskrywing *aggregate* in die Engelse woordelys, sien sy artikel soos volg daar uit:

aggregate

[n] aggregaat; versameling, totaal

[v] aggregeer; versamel, saamvoeg

Alhoewel die kommapunt aandui dat daar by sowel die selfstandige naamwoord as die werkwoord twee verskillende betekenisonderskeidings aangetref word, kry die gebruiker geen konteksleiding nie. Vanuit die oogpunt van teksresepsie is dit miskien voldoende, maar nie vanuit die oogpunt van teksproduksie nie. Soos gesien kan word in die bewerking van *bandwydte* word daar wel in sekere gevalle konteksleiding gegee. Sulke konteksleiding kom egter nie deurgaans by alle inskrywings voor nie en woordeboekgebruikers moet steeds in baie gevalle hulle intuïsie gebruik.

bandwydte bandwidth (*in data communications*); tape width (*e.g. of magnetic tape*)

Die woordeboek bied geen leiding aangaande watter vertaalekwivalent die gebruiklikste is nie. Die vertaalekwivalente word meestal alfabeties gelys (sien *protected location*), maar dit vind ook nie altyd konsekwent plaas nie (sien *root directory*).

protected location beskermde posisie, beskutte posisie

root directory [*n*] stamgids, wortelgids, basisgids

Alhoewel daar in die gebruiksnotas melding gemaak word van die rede vir die verskillende maniere waarop sinonieme gelys word ("Wanneer sinonieme [...] nie alfabeties gelys is nie, het die samestellers doelbewus 'n voorkeur in die doeltaalsinonieme aangedui"), sou die teksproduksiefunksie van die woordeboek heelwat beter gewees het indien die vertaalekwivalente eerder deurgaans konsekwent volgens gebruiksfrekwensie gerangskik is. Daar is 'n hele paar gevalle waar vertaalekwivalente alfabeties gelys is, terwyl daar wel 'n vertaalekwivalent bestaan wat 'n baie hoër gebruiksfrekwensie het as die ander wat daarmee saam vermeld word. Met 'n bietjie meer navorsing kon die samestellers al die vertaalekwivalente volgens gebruiksfrekwensie gelys het wat die woordeboek baie gebruikersvriendeliker sou gemaak het.

Volgens die verduidelikende aantekeninge word woordsoorte in een van twee gevalle aangedui:

- (1) "Wanneer 'n trefwoord dieselfde vorm in verskillende woordsoorte het, word die verskillende vorms aangedui deur 'n woordsoortmerker tussen hakies op afsonderlike nuwe reëls te plaas."
- (2) "Wanneer dit nodig is om die woordsoort van die trefwoord aan te dui, staan die woordsoortmerker tussen die trefwoord en sy sinonieme."

Die aanduiding van verskillende woordsoorte by trefwoorde waar verskillende vorms aangetref word, is heel logies. Wat egter minder sinvol is, is die aanduiding van woordsoorte "wanneer dit nodig is". Die samestellers verduidelik glad nie by watter tipe gevalle so 'n aanduiding nodig sou wees nie. As 'n mens deur die woordeboek blaai, lyk dit asof hierdie aanduiding van woordsoorte lukraak voorkom en nie altyd konsekwent toegepas word nie (sien die woordsoortaanduidings hieronder by *skyn* en *invoegsel*). Dit laat die vraag ontstaan

waarom die samestellers nie maar die woordsoort by al die trefwoorde kon aangedui het nie? Die woordeboek sou beslis daarby gebaat het as die bepaling van woordsoorte nie aan die gebruiker oorgelaat sou gewees het nie.

skyn shine [v]

invoegsel [n] insertion, add-in

Die bewerking van trefwoorde met sinonieme is ook verwarrend. Volgens die verduidelikende aantekeninge word dit soos volg bewerk: "Wanneer die trefwoord [...] 'n sinoniem vir een van die betekenisonderskeidings het, word die vorm sonder die sinoniem eerste gelys." Tog vind hierdie bewerking nie altyd konsekwent plaas nie, want by die aanbieding van *circuit board* word die trefwoord *met* die sinoniem eerste gelys.

circuit board (circuit card) stroombaankaart, stroombaanbord (*waarop stroombane gemonteer is*)

circuit board stroombaanbord, stroombord, kringkaart

As woordeboekgebruiker weet 'n mens nie of jy in sulke gevalle moet aanvaar dat die twee inskrywings verskillende betekenisonderskeidings het nie, want as jy gaan kyk na die vertaalekwivalente wat verskaf word, voel dit eintlik asof hierdie vertaalekwivalente almal sinonieme is.

Die Engelse en Afrikaanse lemmalyste stem glad nie ooreen nie. In 'n poging om via die Afrikaanse lemmalys uit te vind of *stroombord* en *kringkaart* nie dalk wel sinoniem is met *stroombaankaart* nie, blyk dit dat hulle nie eens in die Afrikaanse lemmalys voorkom nie. Alhoewel *kringbord* wel gelys word, word dit nie as vertaalekwivalent aan die Engelse kant gegee nie.

kringbord circuit board, circuit card (*to which circuitry is affixed*)

Nie een van die vertaalekwivalente wat aan die Engelse kant by *circuit board* gelys is (*stroombaankaart, stroombaanbord, stroombord* of *kringkaart*), kom in die Afrikaanse lemmalys voor nie. Sulke inkonsekwentheid maak dit vir die gebruiker byna onmoontlik om by teksproduksie 'n korrekte vertaalekwivalent te kies en bemoeilik ook suksesvolle teksresepisie.

Alles in ag genome blyk dit dus dat die leksikografiese versorging van *Kuberwoordeboek* dikwels gebrekkig en onvoldoende is en daar baie ruimte vir verbetering bestaan. *Kuberwoordeboek* mag dalk 'n leemte vervul in die informasietegnologiebedryf, maar die woordeboek self bevat baie leemtes. Met 'n bietjie meer leksikografiese versorging kon dit 'n baie geslaagder produk gewees het.

Liezl Potgieter
Johannesburg
(liezlp@vodamail.co.za)

Ferdie Weich. *San Dictionary/San-woordeboek. San–Afrikaans–English/English–San–Afrikaans/Afrikaans–San–English.* First edition/eerste uitgawe. 2004, xi + 377 pp. ISBN 0 7978 2704 8. Johannesburg: Macmillan South Africa Publishers (Pty) Ltd. Price/Prys: R198.

The compiler's introduction to this trilingual San–English–Afrikaans dictionary gives important information that helps to place it in perspective. In 1961, the compiler was posted in the territory that is now Namibia doing missionary work until his retirement in 1998. The San community among whom he worked lived, and some continue to live, in Northern Namibia around the Tsumkwe area. According to the compiler and from what is known from other sources, the South African army moved around these Tsumkwe speech communities during the Namibia independence struggle. Because of this association with the South African army during the struggle, the community was resettled at Schmidtsdrift near the Vaal River between Kimberly and Douglas in the Northern Cape. However, this community was not homogenous as it comprised 3 000 !Xuhn and 200 Khwe (a Western Caprivi community, not related linguistically to the !Xuhn). There is no information about the fate of those who remained in the original region of these ethnic communities. However, some sources allege there are some who remained further south of Tsumkwe. As for the Khwe, quite a number of them still live in the Western Caprivi (cf. Takada 2007).

In the introduction, crucial information is furthermore given regarding the identity of !Xuhn, the ethnic language which Weich also describes as Jun|'hoan. !Xuhn is an ethnonym, most probably a dialectal variation of the Jun|'hoan language. Previous and current research on the languages of Northern Namibia places !Xuhn (or !Kung) as the northern variety of Jun|'hoan) (cf. WIMSA 2001). The author seems to have preferred to adopt the ethnic and linguistically problematic label of San, which is purported to denote and provide a broader umbrella term for the Southern African indigenous populations linguistically called Khoisan. However, it should be made clear that San, having no linguistic value, brings together ethnic communities speaking diverse languages. Therefore, this San–English–Afrikaans dictionary is linguistically confusing as it takes an anthropological term for a language. However, there is much linguistic diversity and difference among these communities falling under this label, as is evident from recent research by Güldemann and Vossen (2000) and Snyman (1970, 1974, 1998). From what is currently known, there are at least four language families within what is called Khoisan. Anthropologically, the Khoisan may share certain similarities in their social organization and mode of life. Nevertheless, historically, linguists often make a distinction between Khoi and San (or Non-Khoe). Linguistically, the San are all speech communities belonging to the languages or language families Jun|'hoan, !Xoǝ, †Hoan and N|u, not related to the Khoekhoe group, comprising Nama and Damara (Khoekhoe-gowab), Khwe, Naro, |Gui, ||Gana, Kua, Cua and Shua.

However, in the community development activities of non-governmental organizations such as the Working Group on Indigenous Minorities in Southern Africa (WIMSA) and the Southern Africa San Initiative (SASI), the term San collectively describes all indigenous Khoisan communities as they are characterized by aboriginality and suffering the socio-economic effects of destitution and marginalization. These days, San has therefore become a sociocultural label, and as such, it is not at all suitable for a language or for the name of a dictionary, for that matter. The language treated in this trilingual dictionary should be understood to be !Xuhn, a northern variant of Jun|'hoan. Jun|'hoan is a Northern Non-Khoe language (linguistically San) and it is distantly related to !Xóõ and †Hoan, both spoken in Botswana. This should clarify the question of labels and the linguistic value of ethnic languages, or references thereto in the dictionary, and in this review.

As !Xuhn is related to the Jun|'hoan language, the author correctly associates it with the research carried out by Jan Snyman (1970, 1974, 1998), who was the first to suggest a practical orthography for Jun|'hoan in the early 1960s. The compiler also correctly observes that some researchers such as Prof. E.O.J. Westphal, Mr. W. Zimmermann (of the Language Bureau in Windhoek) and Rev. Flip van der Westhuizen also made significant attempts to develop an orthography for Jun|'hoan. However, it is the work of Patrick Dickens (1994, re-edited 2005), who also compiled a Jun|'hoan dictionary that significantly contributed to the formalization of the orthography. Based on the International Phonetic Alphabet, this writing by Patrick Dickens reversed the former suggested developments that sought to follow the Nguni click type orthography for the !Xuhn language.

Weich's dictionary therefore endeavours to use the Dickens orthography as much as possible. However, it is difficult to judge the extent to which this has been done, especially when considering the elaborate details linguists have presented on Jun|'hoan. There is no mention of the practical orthographic enhancements made to the Dickens orthography by the Nyae Nyae community under the guidance of Amanda Miller, an American linguist who has worked on Jun|'hoan phonetics for a number of years. Without some clarification on and evidence of the application of the phonetics and phonology of the language, it is difficult to tell how the orthography of this dictionary deals with the phonetic and phonological complexity of the language. The section on the consonants and symbols provides evidence of some awareness of the Dickens orthographic principles. However, these explanations and indications cannot be considered adequate for a language with such a varied phonology as !Xuhn. The notes on tone are also very scanty and do not suffice to provide information on the complicated grammar of the language. However, the task here is to assess the published dictionary, which will certainly be used as a reference work even among the speakers themselves.

With its definitions, a dictionary aims to provide such information on the vocabulary of a language that will satisfy the needs of the user. It should ordi-

narily not only familiarize learners with the meanings of lexical items and their correct and idiomatic use, but also provide experts with such linguistic information as will facilitate research and analysis (cf. Haacke 1998). Any dictionary, especially an indigenous language dictionary, should therefore present a compilation that will adequately reflect the grammatical information a lexical item contains in that language. This perspective will serve as guidance in reviewing this dictionary.

As stated earlier, the dictionary is trilingual, providing entries in !Xuhn–Afrikaans–English, followed by English–!Xuhn–Afrikaans, and lastly by Afrikaans–!Xuhn–English. Strictly speaking, these sections of the dictionary present vocabulary lists for the three languages. This has value for a very preliminary field-vocabulary list. Nevertheless, even as field-vocabulary list, this presentation is still inadequate because there is no grammatical sketch of the language. Without the use of example sentences in !Xuhn, it is difficult to appreciate the semantic equivalence of what is provided as Afrikaans or English translations. For a dictionary, this is rather insufficient.

Furthermore, it would have been expected that, since the author indicated that the language is tonal, the dictionary entries would show tone marking to distinguish between homophones and homomorphs, of which there are many in !Xuhn. Linguistically this is important as in many tone languages there will be orthographic homographs which nonetheless have tonal variations. Nowhere in the dictionary, however, this important aspect of the language is catered for. For other researchers and even for the speakers of the language, these vocabulary lists will present daunting difficulties and may be of little value for a grammatical and lexical appreciation of !Xuhn.

As mentioned earlier, the three language sections of the dictionary consist of rudimentary vocabulary compilations. That is, the entry of vocabulary items according to the three languages does not present general linguistic information that would qualify this compilation as a dictionary. There is no lexicographical lemmatization of entries. Headword entries, listed alphabetically, are presented with minimal lexicographical information. The only lexical distinction made is two grammatical categories — verbs and nouns. Any other categories can only be deduced from the provided Afrikaans or English equivalents. The treatment of semantic relations (synonymy, polysemy and hyponymy) cannot be shown in such a rudimentary vocabulary listing. This adds a critical problem to this dictionary: how to identify words grammatically (semantically, syntactically and lexically); and how else to determine the word without providing a grammatical context.

As there are no definitions or descriptions of the headword entries, the only sources of meaning are the English and Afrikaans equivalents. However, dictionary users, especially bilingual or multilingual ones, know how inadequate semantic minimalism can be. There is never any perfect word equivalence between languages. Languages that do not share the same cognitive linguistic structure and cultural knowledge will not necessarily entail similar

things and/or experiences. Normally providing translated sentences that contain the headword assists to understand its meaning, semantically and also the likely syntactic and contextual peculiarities its usage presents. Without definitions, it is difficult to appreciate the value of a vocabulary equivalence for all the languages, but importantly for !Xuhn. In addition, it is difficult to determine the usage in !Xuhn of other grammatical categories such as adverbs and adjectives, and also other aspects of the grammar that would make the target language an object of linguistic analysis (cf. Haacke 1998). As !Xuhn is meant to be the target language, there is nothing that helps the dictionary to realize this objective. !Xuhn words are accorded the same lexicographical treatment as the English and Afrikaans words.

As !Xuhn is basically the target language, its words should therefore be accorded a proper lexicographical treatment to enhance the linguistic value of the dictionary. There is no trilingual balance in the dictionary. Lexicographically it is inadequate, and linguistically it would pose challenges to the speakers of the language. Without showing the words in sentential context, it is difficult to check and determine semantic domains and the morphophonological integrity of the !Xuhn words appearing in the compilation. The lack of tonal notations is another omission. !Xuhn is a tonal language where orthographic homomorphs can bear various tonal presentations. Without tonal notations, it would therefore be difficult to ensure that the recorded entry lexically and semantically fulfils its linguistic role. This compilation which is merely a word-list will need more work to enable it to contribute to the study and development of this language, which is among the Khoisan languages threatened with extinction.

The Khoisan languages of Southern Africa are underresearched and their sociolinguistic status is precarious (cf. Batibo and Smieja 2000). Any text or vocabulary produced on them should add value to the endeavour to record them and to contribute to their preservation and promotion (cf. Haacke 1998). As small community languages, under constant risk of being assimilated by languages of greater communication, it is important also to produce lists that will assist in easily identifying vocabulary matches with other languages (cf. Haacke 1998).

There are several issues that should be paid attention to if this dictionary will have a second edition:

A close look at the treatment of dictionary entries suggests that the original compilation was based on Afrikaans. The English is simply a translation of the Afrikaans working language. There is nothing wrong with this. However, due care and attention should be paid to the lexical generation of the target language. In this dictionary, it will be observed, however, that in many cases where !Xuhn would present polysemy or synonymy, these lexical aspects have not been captured in the target language but in the working languages translating the various terms. English and Afrikaans were introduced to provide access to the target language, but the user is often left in the dark. A dictionary

should elucidate the words of the target language by their translation into other languages (cf. Kilian-Hatz 2003).

Lexical and orthographic harmonization is important for a dictionary. A dictionary such as this one can have the negative effect of conventionalizing lexical and orthographic practices that might not have been adopted by related speech communities. Even though there is a visible attempt to follow the Dickens orthography in this dictionary, there were many other enhancements to the Dickens orthography added in the past decade by other linguists who worked with the Nyae-Nyae community speaking the Jun|'hoan language (cf. WIMSA 2001). Indeed, the linguistic issues the compiler alludes to in the introduction will require expert linguists not only in phonetics and phonology, but also in semantics and syntax to adequately address the problems arising in the use of this compilation. There is no information on the technical aspect of the dictionary database. However, it looks as if the compilation was made from a document file. Modern lexical undertakings benefit a great deal from the use of a dictionary database with interactive management software. This technology can be handy in the practical and linguistic manipulating of the dictionary entries.

A lexicographical undertaking can be a personal and sentimental matter, affording the compiler to readily access the vocabulary he has recorded of the language of the people with whom he daily interacts. This seems to be the evident albeit undeclared objective of this dictionary. Therefore it is more than just a published personal memory prompt. The credit for this publication is the testimony of the compiler's love for and dedication to the language of the speakers with whom he lived for 30 years. This is commendable. However, within this long period, the author must have acquired sufficient knowledge of the language to have done more to make it linguistically valuable for its speakers and researchers. In fact, a publication such as this entails sharing the fruits of personal labour and knowledge with the users. The energy spent on these languages, which are increasingly assailed by sociolinguistic dynamics and modernity, demands that dictionaries should contribute to ensure their functionality of use.

References

- Batibo, Herman M. and Birgit Smieja.** 2000. *Botswana: The Future of Minority Languages*. Duisberg Papers on Research in Language and Culture 40. Frankfurt/New York: Peter Lang.
- Dickens, Patrick.** 1994. *English–Ju|'hoan, Ju|'hoan–English Dictionary*. Quellen zur Khoisan-Forschung 8. Cologne: Rüdiger Köppe Verlag.
- Dickens, Patrick J.** 2005. *A Concise Grammar of Ju|'hoan: With Ju|'hoan–English Glossary and Subject Index*. Research in Khoisan Studies 17. Cologne: Rüdiger Köppe Verlag.
- Güldemann, Tom and Rainer Vossen.** 2000. Khoisan. Heine, Bernd and Derek Nurse (Eds.). 2000. *African Languages: An Introduction*: 99-122. Cambridge: Cambridge University Press.

- Haacke, Wilfrid.** 1998. A Khoekhoe Dictionary in the Making: Some Lexicographic Considerations. Schladt, Mathias (Ed.). 1998. *Language, Identity and Conceptualization among the Khoisan*: 35-64. Research in Khoisan Studies 15. Cologne: Rüdiger Köppe Verlag.
- Kilian-Hatz, Christa.** 2003. *Khwe Dictionary. With a Supplement on Khwe Place Names of West Caprivi by Matthias Brenzinger*. Namibia African Studies 7. Cologne: Rüdiger Köppe Verlag.
- Köhler, Oswin R.A.** 1981. Les langues Khoisan. Perrot, Jean (Ed.). 1981. *Les langues dans le monde ancien et moderne. Volume 1: Les langues de l'Afrique subsaharienne — Pidgins et créoles*: 455-615. Paris: Éditions du C.N.R.S.
- Snyman, J.W.** 1970. *An Introduction to the !Xū Language*. Cape Town: A.A. Balkema.
- Snyman, J.W.** 1974. *Žu|hōasi: Fonologie en Woordeboek*. Cape Town: A.A. Balkema.
- Snyman, J.W.** 1998. An Official Orthography for Žu|hōasi Kokx'oi. Schaldt, M. (Ed.). 1998. *Language Identity and Conceptualization among the Khoisan*: 95-136. Research in Khoisan Studies 15. Cologne: Rüdiger Köppe Verlag.
- Takada, Akira.** 2007. Changing Locality and Ethnic Cohesion among San in Ohangwena, Namibia. Fujioka, Yuichiro and Masashi Iida (Eds.). 2007. *ASAFAS Special Paper 9*: 75-90. Kyoto: Graduate School of Asian and Africa Area Studies (ASAFAS).
- Working Group of Indigenous Minorities of Southern Africa (WIMSA).** 2001. *The Penduka Declaration on the Standardisation of Ju and Khoe Languages. 20–22 April 2001*. Windhoek, Namibia: Penduka Training Centre.

Andy Chebanne
Department of African Languages
University of Botswana
Gaborone
Botswana
(chebanne@mopipi.ub.bw)

Herbert Ernst Wiegand. *Internationale Bibliographie zur germanistischen Lexikographie und Wörterbuchforschung. Mit Berücksichtigung anglistischer, nordistischer, romanistischer, slavistischer und weiterer metalexikographischer Forschungen.* Volumes 1 and 2. 2006, 1458 pp. Volume 1: ISBN-13: 978-3-11-013758-3/ISBN-10: 3-11-013758-5. Price: €198. Volume 2: ISBN-13: 978-3-11-019026-7/ISBN-10: 3-11-019026-5. Price: €198. Berlin: Walter de Gruyter.

These are the first two volumes of a three-part international bibliography of German lexicography and dictionary research. The third part will appear in two separate volumes in the near future. This review therefore does not deal with the entire publication, but mainly highlights Wiegand's approach to the compilation of this bibliography, discussing several aspects of the items he included.

The title suggests that "international" literature was considered for inclusion, which means that not only research material dealing with German lexicography will be found in this bibliography. The titles include all aspects of lexicography, written in more than 15 languages and treating the lexicography of many more languages, from its earliest stages up to the present.

Very few of the existent bibliographies in other fields give a comprehensive account of the compilation approach and process, nor do they provide detailed users' guidelines. To refer to only a few examples: Jessen's *Bibliographie der Autobiographien* (1987) comments only very cursorily on the rationale for the bibliography. Other bibliographies include some explanations of the way and manner of compilation, but not very detailed. Examples of this type can be found in Blackwell's *Bibliography of the Philosophy of Science, 1945–1981* (1983: xiii-xvi) and Van Noppen and Hols's *Classified Bibliography of Publications, 1985–1990* (1990: 1-4). Gazdar et al. (1978: ix-xviii) explain their selection process and the sources included, but give no users' guidelines. The same applies to Ferguson's *Bibliography of Translations from Medieval Sources, 1943–1967* (1974: vii-x) and Baldensperger and Friederich's *Bibliography of Comparative Literature* (1950: XV-XVIII). One bibliography which contains more detailed instructions for use in the way Wiegand does, is Fazeka's *Bibliographie [...] der ungarischen Literatur in deutscher Übersetzung, 1774–1999* (1999). Fazeka gives information on the structure of the entries, the types of items used within the entries, and the way these should be interpreted.

Wiegand (2006: VII) states that he collected the items for this bibliography through the years, while working in the field of lexicography. It is an expansion of the "Bibliographie zur Wörterbuchforschung von 1945 bis auf die Gegenwart: 2200 Titel. Ausgewählt aus germanistischer Perspektive", which first appeared in 1988 in the *Studien zur neuhochdeutschen Lexikographie*. The information contained in these volumes cannot be found in quick internet searches, nor can it be collected within a short period of time. This truly is a painstaking task, taking years of discipline and dedication. The selection was done according to the four research areas Wiegand had already identified many years ago:

(a) research into dictionary use, (b) critical dictionary research, (c) historical dictionary research and (d) systematic dictionary research.

Following on the preface, Wiegand's users' guidelines are comprehensive, especially when compared to those of the other bibliographical works mentioned above. He explains in great detail how the bibliography is structured, and what types of information are included. The users' guidelines are followed by an introduction, in which Wiegand carefully expounds his selection criteria. In addition, he promises to add an epilogue ("Nachwort") in the last volume of the bibliography, in which, among others, he will give more information on the compilation of a useful subject index. Apart from the bibliographical list itself, there will be three indices, namely, (a) the index containing a list of all the excerpted textbook and handbook volumes, (b) an index of author names, and (c) a subject index.

According to Wiegand (2006: XXI), this bibliography is polyaccessible, which means that it is possible to access data in four different ways. This will, of course, only be fully possible once all the volumes have been published. In addition to access via the alphabetical list of entries, all three indices can also be used when looking for a particular title. The alphabetical subject index, for example, will list all the occurrences of a particular expression in a title, as well as the frequency of its occurrence. For example, in the item

Bedeutungsparaphrase 282, 476, 803, [...], 23870; **66**

the last number (in bold print) gives the frequency with which the expression is found in the bibliography. The subject index also contains cross-references to synonyms and other related index entries. Any additional information, found in the expanded entries as discussed below, are furthermore accounted for in the subject index.

In the users' guidelines, Wiegand explains the types of bibliographical entries included. Firstly, there is the simple bibliographical entry, with the following appearance:

136 Agricola, Erhard: Darstellung des semantischen Systems in Wörterbüchern. In: Sprachwissenschaftliche Informationen 6. 1983, 74-78.

This type of entry always consists of two parts, namely, the title number (here **136**), and the title itself.

Secondly, there are expanded bibliographical entries, which consist of various parts which may differ in character. Sometimes, a title in its entirety is mentioned in an entry, which means that the user does not have to make any additional cross-references. This can be called an "independent" bibliographical entry. If the title is only mentioned in an abbreviated manner, the entry is a so-called "dependent" or "incomplete" entry. The following can serve as an example:

144 Ahadi, Sharham: Zum Stand der Deutsch-Persischen Lexikographie. In: Studien zur zweisprachigen Lexikographie mit Deutsch VI [...] ↑, 27-66.

Dependent entries only occur when the titles appear in handbook or reader volumes, and they can be recognized by the presence of the marks of omission "[...]", as well as by the cross-reference arrow which follows them. This means that part of the title of the volume has been left out, and that the particular volume is indicated by the cross-reference arrow. If a user wants to know in which volume the title in entry number 144 was published, he/she should cross-refer to the letter S in the list and look for the complete title under "Studien zur zweisprachigen Lexikographie mit Deutsch VI [...] 2001". This will lead to the following entry in volume 3:

20735 Studien zur zweisprachigen Lexikographie mit Deutsch VI. Hrsg. v. Herbert Ernst Wiegand, Hildesheim [etc.] 2001 (Germanistische Linguistik 163/2001). [Daraus: 144, 11259, 13356, 13407, 14506, 14690].

The title number 144 appears in square brackets, which proves that the user has found the right cross-reference. The six numbers appearing after "Daraus" indicate which titles from the particular volume have been included in the bibliography.

Wiegand also sometimes gives additional information in expanded entries, which may look like this:

889 Atkins, Beryl T. Sue/Hélène Lewis/Della Summers/Janet Whitcut: A Research Project into the Use of Learners' Dictionaries. In: *The Dictionary and the Language Learner* [...] 1987[↑], 29-43.
Vgl. *Lex. Today*, 16 · Vgl. Dolezal/McGreary 1996, Nr. 8 · Vgl. *Ped. Lex. Today*, 6.

The different parts of the expanded entry have several reference functions. For example, references to other publications can be made, e.g. to advertisements of the particular publication, to a particular edition of the publication, to the author, to a bibliography, to special-field dictionary articles, glossaries, textbook articles, annotated glossaries, short reviews, microfilms, obituaries, press texts, project descriptions, reviews, review articles, conference proceedings, publishers' catalogues, and reports from the publishing house. In the following example, it is shown that the publication has a summary in another language:

882 Atkins, Beryl T. Sue/Krista Varantola: Monitoring Dictionary Use. In: *Using Dictionaries* [...] 1998[↑], 93-11. [franz. Résumé 206; dt. Zusammenfassung 210].
Vgl. *Ped. Lex. Today*, 5.

Even abbreviations of titles are explained, which happens in square brackets. For example, entry number 190 contains the explanation of DOST = *Dictionary of the Older Scottish Tongue*. Indications of where a particular publication first appeared can also be found in square brackets.

The bulk of the titles in the bibliography, of course, deal with German lexicography, and are written in German. But also found are many titles dealing with the lexicography of, for example, Ancient Greek (3655), Albanian (5518), Arabic (4108-4115, 5491-5492, 5494, 10374), Aramic (10505), Armenian (4907), Australian aboriginal languages (964, 4526-4530, 11061), Belorussian

(13580), Bulgarian (1247, 10143-10148), Caribbean English (352-353), Catalanian (2380-2381), Caucasian languages (15231), Central-American languages (1663, 1768), Chinese (1536, 3492, 3721, 11884), Indian Ocean Creole (4087), Croatian (6297), Czech (1200, 10865), Danish (5787), Dutch (1184), Esperanto (2357-2367), Estonian (12110-12111, 15488), Flemish (758), French (420), Frisian (761, 3156-3157, 4845) and Old-Frisian (2913-2918, 3143-3144), languages of the Indian subcontinent (1197, 10494-10497), Indonesian languages (1455), Irish (15483-15488), Hebrew (1385), Hettite (10305), Hungarian (1334-1336, 1714), Icelandic (2065, 10128-10134), Iranian (8295), Italian (1206-1207), Japanese (3492), Latvian (1269), Malaysian (1802), Norwegian (363), Persian (734, 1753, 17790), Philippine languages (2337), Polish (717, 10373), Portuguese (2194-2197, 16448), Romanian (995-997, 3731-3732), Russian (21), Sanskrit (6902-6903, 11733, 11746), Scottish (3573), Serbo-Croatian (1799), Slovenian (147, 1199), Spanish (22-24), Swedish (336 and 345), Sumerian and Akkadian (3805, 4549), Tamil (9966-9968, 10166), Turkish (4947, 11822), Vietnamese (15227), as well as aspects of the lexicography of the North American Indian languages (17282-17283), Andalusian lexicography (17564), Pakistani lexicography (9865), Mexican (12043) and Chilean lexicography (13722), Tibeto-Burman (13457) and Burmese (15586) lexicography, etc.

As can be expected, the lexicography of different German dialects are covered in detail. For example, represented are Prussian dialects (738), Luxemburgian German (5568), the dialects of Hamburg and Bremen (12059-12067), the dialects of the Germans in Transylvania (965, 7983-7985, 8009, 9874, 10584, 12659, 16642-16645), Swiss German (1527, 2635, 4341-4351, 5504, 7996-7997, 12054), Austrian German (1584, 6480-6482), and even the dialect of Beatenberg, a small town in the Thun lake region (3112).

Furthermore, there are many titles pertaining to special-field lexicography. Subject-field lexicography in general is represented, e.g. by Bergenholtz and Schaefer (1912), Bergenholtz and Tarp (1915-1916), Hupka (9711-9716) and Kalverkämper (10296-10300). Many large volumes dealing with special-field lexicography are included (5778-5783). Special-field studies include those on medicine (150, 12998, the latter being a record of 16th century medical terminology), music dictionaries (1286-1289, 3902-3903, 4033-4034, 5391-5399), dictionaries of literature (1621), law dictionaries (2383-2385), glossaries of video-making technique (2542-2543), ESP lexicography in Jordan (4721-4725), Biblical and theological lexicography (5482, 6381-6386, 6906-6908), a special-field dialect dictionary (5788), a dictionary on philosophy (6841), dictionaries for the blind (12127), the special-field language of lexicography (12647), Chinese special-field lexicography (14227), and even alchemy (17564). Etymological studies, such as Lindner's on the lexicon of Indo-Germanic verbs (12608), are mentioned. Terminology in general is represented by 793, 5109-5110, and 5294-5297. Some works with recommendations for the selection, formation and definitions of technical terms (BSI BS 3369, entry 3093) are included. Also listed are Gerhard Budin's articles on terminological data banks (3124-3126), Finnish experts' writ-

ings on terminography (10288-10290), as well as Mariëtta Alberts's contributions (240-252).

Bilingual lexicography is represented, for example by works on the language pairs German–Persian (1674), German–Indian (3654), Chinese–English (4205-4215, 12534), Japanese–German (5746-5750, 8297), Russian–English (5847), English–Japanese (8351, 8519, 10807), Finnish–German (10372), German–Hungarian (10809), Danish–German (12120-12122), Dutch–Hungarian (14396), German–Greek (15835), and many more. Even works dealing with bilingual lexicography concerning Bantu and Indo-European languages (13605) are included.

Computer lexicography also receives attention. For example, titles about specific computer-assisted projects are mentioned, as one on Alemannic dialectology in the computer era (283), and one on dictionaries on the internet (1769). In entries 3165-3168 Thomas Burch and others write about Middle High German dictionaries available on CD-ROM; 3676 deals with hypertext and the production of digital dictionaries; other works are on computational linguistics (4283), the digitalization of dictionaries (4841-4844), the electronic Duden (5501-5502), the Heidelberg Hypertext Server (8599-8600), even though some of these articles may already be outdated.

References to the use of dictionaries can be found, e.g. the article by Beattie (1596) dealing with the teaching of dictionary use, or the article of Becker (1626) treating the use of an etymological dictionary as a teaching device. Entry 2308 deals with language learners' use of bilingual dictionaries, and 2316 with an electronic bilingual Icelandic dictionary.

Corpus lexicography (1539, 6001-6004, 16225) is taken into account, as well as linguistic aspects of lexicography such as the use of quantifiers in dictionaries (3155). Phonological, morphological and syntactic issues in dictionaries are treated by Caluwe and others (entries 3377-3378). Socio-linguistic articles are included, for example entry 4515 treats vulgarities in lexicography. Other topics deal with the handling of figurative senses in learner's dictionaries (4538) and the use of "frames" in dictionaries (3094).

Many of the well-known leading personalities in lexicography are listed with all their publications, such as B.T. Sue Atkins (864-890), Paul Bogaards (2471-2502), Nicoletta Calzolari and her colleagues (3384-3409), Anthony P. Cowie (4159-4185), Helmut Felber, who has written much on terminological matters (5870-5888), William Frawley (6308-6321), Frieberthäuser who is known for his work on dialect lexicography (6347-6366), Kurt Gärtner (6644-6670), Dirk Geeraerts who wrote on lexicography and linguistic matters (6743-6757), Joachim Göschel, who is also noted for his work on dialect lexicography (7204-7231), Rufus Gouws (7304-7370), Philip B. Gove (7371-7392), Günther Haensch (7894-7939), Rainer Hartmann (8181-8283), Ulrike Haß-Zumkehr (8306-8337), Franz Joseph Hausmann (8385-8505), Ulrich Heid (8568-8597), Helmut Henne (8751-8792), Thomas Herbst (8877-8904), Lothar Hoffmann (9257-9267), Werner Hüllen (9634-9660), Robert Ilson (9803-9824), Howard Jackson (9913-9921), Matthias Kammerer (10307-10339), Dieter Karch (10397-10409), Günther Kempke

(10589-10614), Alan Kirkness (10755-10791), Ruth Klappenbach (10837-10849), Friedrich Kluge (10955-10992), Francis Knowles (11043-11055), Hans-Peder Kromann (11541-11564), Peter Kühn (11667-11723), Hans Küpper (11648, 11768-11786), William Kurrelmeyer (11792-11810), Sidney Landau (11930-11942), Theodor Lewandowski (12442-12452), Wolfgang Lindow (12610-12619), Helene Malige-Klappenbach (13101-13119), Yakov Malkiel (13122-13135), Sven-Göran Malmgren (13138-13153), Igor A. Mel'čuk (13855-13879), Wolfgang Mentrup (13932-13958), Dieter Möhn (14328-14355), Noel Edward Osselton (15679-15703), Max Pfister (16127-16166), Rudolf Post (16453-16472), Danie Prinsloo (16578-16593), Ulrich Püschel (16693-16714), Allen Walker Read (16947-16962), Oskar Reichmann (17014-17067), Kurt Rein (17094-17118), Alain Rey (17207-17246), Josette Rey-Debove (17247-17269), Fred Riggs (17372-17377), and Martha Ripfel (17388-17396).

Many doctoral dissertations and other studies found their way into the bibliography. For example, Benson's on Asian varieties of English and their treatment in dictionaries (1803), Beuke's on the lexicographical treatment of Namibian Afrikaans (2154), Brand's on the lexicographical inconsistency in the central list of the *Major Dictionary/Groot Woordeboek* (2791), and Keyser's research into training for lexicographers (10663).

When using the bibliography from the perspective of Africa and Southern Africa, entries on a variety of studies on African and Southern African lexicography are found. Many articles are listed from *Lexikos*, the official journal of the African Association for Lexicography (AFRILEX). Aspects of lexicography in Africa are relatively well covered. For example, included are the volume *Lexicography in Africa* dealing mainly with Western Africa (1000), publications of Branford (2813-2815), Busane's article on lexicography in Central Africa (3271), Bwenge's work (3328-3330), Chimhundu's article on the ALLEX project (3723), Francophone lexicography in Africa (10291), Madiba and others' work (13039), the research of Mann (13182, 13184), Mbiti's study on African proverbs (13604), an old article on Bantu languages (13831), Moropa and Kruger's work on mistranslation in Kropf's dictionary (14498), Mtintsilana's work (14568), and writings of the Pan South African Language Board (15812-15818).

Afrikaans lexicographical studies can be found, e.g. by Herman Beyer (2158-2161), Willem Botha (2666-2678), Adelia Carstens (3500-3509), Pieter Harteveld (8154-8158) and D.C. Hauptfleisch (8372), as well as those published by the *Woordeboek van die Afrikaanse Taal* (WAT) (3192-3199, 3248). Shona lexicography is represented by Chabata (3647-3651) and also by Chitauro-Mawena (3729), Mavhu (13545) and Mpofu (14562-14563). Studies on Swahili lexicography by Chuwa (3785) and others (6296) are mentioned. Further entries deal with lexicography in Gabon (13546-13551, 15472) and the Congo (5528-5534, 13606-13607), and with Somali (7602), Khoisan (8035), Kiswahili (9866), Ndebele (10676-10677, 13552) and Xitsonga (13456). IsiXhosa is represented by Mfaha (14129), Mini (14228-14229) and Mtuze (14569), Northern Sotho by Mojela (14373-14374) and Mphahlele (14559-14561), Setswana by Mokone (14378), isi-

Zulu by Mpungose (14564), Nkabinde (15362-15364) and Pienaar (16223-16224), to mention only a few.

If interesting lexicographical topics are looked for, they can be found in abundance. Studies on Martin Luther's vocabulary (1675, 12408) are juxtaposed with studies of the internet as a source for linguistic variations (2192). Subjects range from a dictionary of Shakespeare's informal English (2338), through Egyptian prosopography according to Demotic sources (3078), to Byzantine (3338) and Catholic (3611) encyclopedias and dictionaries. Several articles dealing with the history of Dutch lexicography (3809-3853) can be found, even a very specialised topic such as the treatment of the word *negre* in dictionaries of the *ancien régime* (4536), a bio-bibliographical work on German linguistics of the 18th century (5671, 5769), and an article on searching for vowelless words (6276). A study on children's vocabulary (6294) is included. There is a report on a Slovene–English false friends dictionary (12144), as well as one on the glossary in the Rheims New Testament of 1582 (13544). Some very old titles, useful for historical reasons, are found (such as 10100, 10101-10102), which is the preface to and plan of the English Dictionary compiled by Samuel Johnson between 1747 and 1755. Represented are studies from before World War II on Old High German (10413-10434, 12553-12554), some old Dutch studies (10993-10994), as well as Wallace Lindsay's works on old Latin glossaries (12621-12628).

One question that a user of such a bibliography will ask, is: To what extent can such a work claim to be comprehensive and representative? Only a few items deal with, for example, Ethiopian lexicography, while a large number cover German lexicography. Wiegand (2006: XXV) explains that in the selection of sources for the bibliography, it cannot be determined in advance which titles will be included, and which types of publications will be excluded. There are no fixed rules. One requirement, however, is that the bibliographical details have to be correct. Therefore, representativity is perhaps not the most important criterion with which to judge this publication. Everything considered, it is very comprehensive, in spite of the fact that these titles were documented "in passing" over a long period of time. The collection is large and unique, and very informative.

Wiegand (2006: XXVI) claims that he deliberately omitted details of the publisher. Sometimes, however, names of the publishers as well as the price of the book and even the ISBN number slipped in. This happens, for example, in entries 10477-10483, 11733, 12654, 12739, 14184-14185, 15745.

Although some minor typing errors occur in German (entry 4584), in English (entries 973, 1068, 4751, 12928), and a few in French titles (entries 3061, 6377), it is rather unfortunate that a relatively high percentage of Dutch and especially Afrikaans titles contain typing errors. This happens, for example, for Dutch in entries 118, 2765, 5324, 9013, 9797, 9927, 10006, 11578, 14306 and even more for Afrikaans in entries 2155, 2164-2165, 3193, 3199, 3503, 3506, 3892, 4589, 5127, 5487-5488, 5490, 5702, 5862, 7310, 7313-7314, 7331-7332, 7335, 8154-8155, 8157, 14541-14542, 15526, 15745, 16580).

In spite of its beautiful appearance, it is therefore actually somewhat frustrating that this is a printed publication. It is an expensive resource, and many of the typing problems listed above could have been corrected immediately if this bibliography had been published online. In addition, an online resource such as this could frequently be updated, which would increase its usefulness. Nevertheless, because of its comprehensiveness, the usefulness of this bibliography is beyond any doubt, and its lifespan will be long. It can be used fruitfully in searches for literature on lexicography, and it is therefore recommended that every library and especially university libraries, should have a copy.

References

- Baldensperger, F. and W.P. Friederich.** 1950. *Bibliography of Comparative Literature*. New York: Russell & Russell.
- Blackwell, R.J.** 1983. *A Bibliography of the Philosophy of Science, 1945–1981*. Westport, Connecticut: Greenwood Press.
- Fazekas, T.** 1999. *Bibliographie der in selbständigen Bänden erschienenen Werke der ungarischen Literatur in deutscher Übersetzung (1774–1999)*. Hamburg: Eigenverlag des Verfassers.
- Ferguson, M.A.H.** 1974. *Bibliography of English Translations from Medieval Sources, 1943–1967*. New York: Columbia University Press.
- Gazdar, G., E. Klein and G.K. Pullum.** 1978. *A Bibliography of Contemporary Linguistic Research*. New York: Garland.
- Jessen, J.** 1987. *Bibliographie der Autobiographen. Band 1*. Munich: Saur.
- Van Noppen, J.-P. and E. Hols.** 1990. *Metaphor II. A Classified Bibliography of Publications, 1985–1990*. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, H.E.** 1988. *Bibliographie zur Wörterbuchforschung von 1945 bis auf die Gegenwart: 2200 Titel. Ausgewählt aus germanistischer Perspektive*. Wiegand, H.E. (Ed.). 1988. *Studien zur neuhochdeutschen Lexikographie* 6(2): 627–821. Germanistische Linguistik 87–90. Hildesheim/Zurich/New York: Georg Olms.

Maria Smit
Department of Music
University of Stellenbosch
Stellenbosch
Republic of South Africa
(msmit@sun.ac.za)

Publikasieaankondigings / Publication Announcements

- B.T. Sue Atkins (Editor). *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. 1998, 214 pp. ISBN 3-484-30988-1. Lexicographica. Series Maior 88. Tübingen: Max Niemeyer. Price: €58. (See review article in this issue.)
- W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Deel XII*. 2005, xxiii + 760 pp. ISBN 0 9584401 8 2 (plastiekband), 0 9584401 9 0 (leerband). Stellenbosch: Buro van die WAT. Prys: R290 (plastiekband), R580 (leerband). (Resensieartikel in hierdie uitgawe.)
- Albert Busch. *Diskurslexikologie und Sprachgeschichte der Computertechnologie*. 2004, XIV + 478 pp. ISBN 3-484-31252-1 (Pb.). Reihe Germanistische Linguistik 252. Tübingen: Max Niemeyer. Preis: €138. (Rezension in dieser Nummer.)
- Julie Coleman. *A History of Cant and Slang Dictionaries. Volume I: 1567–1785*. 2004, xii + 259 pp. ISBN 0 19 925470 0 (Hb.). Oxford: Oxford University Press. Price: £45. (See review in this issue.)
- Julie Coleman. *A History of Cant and Slang Dictionaries. Volume II: 1785–1858*. 2004, xiv + 338 pp. ISBN 0 19 925470 2 (Hb.). Oxford: Oxford University Press. Price: £45. (See review 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.
- Philippe Humblé. *Dictionaries and Language Learners*. 2001, 211 pp. ISBN 3-89846-110-6. Frankfurt am Main: Haag und Herchen. Price: €23. (See review article in this issue.)
- Christa Kilian-Hatz. *Khwe Dictionary. With a Supplement on Khwe Place-names of West Caprivi* by Matthias Brenzinger. 2003, x + 431 pp. ISBN 10 3-89645-083-2. Namibian African Studies 7. Cologne: Rüdiger Köppe Verlag. Price: €52.80. (Review in this issue.)
- Phillip Louw (Senior Redakteur/Senior Editor). *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. 2007, xii + 578 pp. ISBN

978 0 19 578742 9. Kaapstad/Cape Town: Oxford University Press Southern Africa (Pty) Ltd. Prys/Price: R99.95. (Resensie in hierdie uitgawe.)

Jack Lynch and Anne McDermott (Editors). *Anniversary Essays on Johnson's Dictionary*. 2005, xi + 245 pp. ISBN-13: 978 0 521 84844 2 (Hb.). ISBN-10: 0 521 84844 X (Hb.). Cambridge: Cambridge University Press. Price: £45.

M.O. Mbatha (Editor). *Isichazamazwi SesiZulu*. 2006, vi + 1353 pp. ISBN-13: 978-1-919959-84-9. ISBN-10: 1-919959-84-x. Pietermaritzburg: New Dawn Publishers. Price: R450 (Hb.), R250 (Sb.). (Review in this issue.)

Hilary Nesi. *The Use and Abuse of EFL Dictionaries: How Learners of English as a Foreign Language Read and Interpret Dictionary Entries*. 2000, VI + 156 pp. ISBN 3-484-30998-9. Lexicographica. Series Maior 98. Tübingen: Max Niemeyer. Price: €54. (See review article in this issue.)

Anton F. Prinsloo. *Sleng: Woorde, uitdrukkings en hul herkoms*. 2006, vii + 256 pp. ISBN 978 186891 813 3 (1 86891 813 0). Kaapstad: Pearson Education South Africa (Pty) Ltd. Prys: R185. (Resensie in hierdie uitgawe.)

Allen Reddick (Editor). *Samuel Johnson's Unpublished Revisions to the Dictionary of the English Language*. 2005, xxix + 425 pp. ISBN-13: 978 0521 84470 3 (Hb.). ISBN-10: 0 521 84470 3 (Hb.). Cambridge: Cambridge University Press. Price: £100.

P.M. Sibula. *IiFakhalthi, amaSebe namaCandelo eMfundo/Faculties, Departments and Academic Divisions/Fakulteite, Departemente en Akademie Afdelings. ULawulo neeNkonzo/Administration and Services/Administrasie en Dienste*. 2004, 13 pp. Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/ Universiteit Stellenbosch Taalsentrum. <http://www.sun.ac.za/taalsentrum>.

P.M. Sibula & C.S. Basson. *Isigama sobuNgcali beMfundo/Generic Academic Terminology/Generiese Akademie Terminologie*. 2004, 28 pp. Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.

P.M. Sibula & C.S. Basson. *Isigama sobuNgcali boLawulo/Generic Administrative Terminology/Generiese Administratiewe Terminologie*. 2004, 31 pp. Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.

Pumlani Sibula. *Isigama sobuNtlalontle/Social Work Terminology/Maatskaplike-werk-Terminologie*. 2006, vi + 42 pp. ISBN 0-9802503-2-3. Stellenbosch: IZiko

leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/ Universiteit Stellenbosch Taalsentrum.

Pumlani Sibula & Erina Basson. *Isigama seSosiyoloji/Sociology Terminology/Sosiologie-Terminologie*. 2006, iii + 36 pp. ISBN 0-9802503-3-1. Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.

Pumlani Sibula & Deirdré Roos. *Isigama esingengabula-zigcawu kwiSayikholoji/An Introduction to Psychology Terminology/'n Inleiding tot Sielkunde-Terminologie*. 2006, viii + 70 pp. ISBN 0-9802503-4-X. Stellenbosch: IZiko leeLwimi le Univesithi yaseStellenbosch/Stellenbosch University Language Centre/Universiteit Stellenbosch Taalsentrum.

Chan Sin-wai. *Translation and Bilingual Dictionaries*. 2004, iv + 190 pp. ISBN 3-484-39119-7 (Pb.). Lexicographica. Series Maior 119. Tübingen: Max Niemeyer. Price: €56. (See review article in this issue.)

A.J. Smit (Hoofredakteur), W.A. Landman, A.E. Gerber en P.L. de Jager (Redaksionele Komitee). *Opvoedkundewoordeboek. Vertalend, Verklarend. Afrikaans/Engels. Deel 1: A–G*. 2007, viii + 388 pp. Selfuitgawe: A.J. Smit, Trevorstraat 33, Murrayfield 0184.

Włodzimierz Sobkowiak: *Phonetics of EFL Dictionary Definitions*. 2006, 249 pp. ISBN 83-7177-450-8. Anglistyka. Poznań: Wydawnictwo Poznańskie. Price: 38 zł. (Review in this issue.)

Birgit Steinbügl: *Deutsch–englische Kollokationen: Erfassung in zweisprachigen Wörterbüchern und Grenzen der korpusbasierten Analyse*. 2005, VI + 189 pp. ISBN 3-484-39126-X. Lexicographica. Series Maior 126. Tübingen: Max Niemeyer. Price: €98. (Review in this issue.)

Renata Szczepaniak: *The Role of Dictionary Use in the Comprehension of Idiom Variants*. 2006, VII + 158 pp. ISBN-13: 978-3-484-39131-4. ISBN-10: 3-484-39131-6. Lexicographica. Series Maior 131. Tübingen: Max Niemeyer. Price €104.

P.J. Taljaard. *Tweetalige voorsetselwoordeboek*. Nuwe uitgawe. 2003, 420 pp. Pretoria: Content Solutions Online. Prys: slegs CD: R150; slegs boek: R200; CD met boek: R225.

Jenny Thumb. *Dictionary Look-up Strategies and the Bilingualised Learner's Dictionary: A Think-aloud Study*. 2004, XII + 127 pp. ISBN 3-484-39117-0. Lexicographica. Series Maior 117. Tübingen: Max Niemeyer. Price: €48. (See review article in this issue.)

- Yukio Tono. *Research on Dictionary Use in the Context of Foreign Language Learning: Focus on Reading Comprehension*. 2001, xii + 257 pp. ISBN 3-484-39106-5. Lexicographica. Series Maior 106. Tübingen: Max Niemeyer. Price: €74. (See review article in this issue.)
- S.L. Tshabe (Editor-in-Chief), F.M. Shoba (Final Editor). *The Greater Dictionary of isiXhosa. Volume 1: A to J*. 2006, xxxviii + 839 pp. ISBN 0 9499 7485 4. Alice: IsiXhosa National Lexicography Unit, University of Fort Hare.
- G.J. van Wyk (Redakteur), A.E. Cloete, H.C. Liebenberg, H.J. Lubbe en A. Prinsloo. *Etimologiewoordeboek van Afrikaans. Supplement*. 2007, xiii + 213 pp. ISBN 0-9584971-4-1. Stellenbosch: Buro van die WAT. Prys: R150.
- H.C. Viljoen (Redakteur/Editor). In medewerking met/in collaboration with N.F. du Plooy en/and S. Murray. *Kuberwoordeboek Afrikaans–Engels, Engels–Afrikaans/Cyber Dictionary Afrikaans–English, English–Afrikaans*. Onder beskerming van/Under the auspices of Vriende van Afrikaans. 2006, 518 pp. ISBN-13: 978-1-86919-133-7. ISBN-10: 1-86919-133-1. Pretoria: Protea Boekhuis. Prys/Price: R250. (Resensie in hierdie uitgawe.)
- Ferdie Weich. *San Dictionary/San-woordeboek. San–Afrikaans–English/English–San–Afrikaans/Afrikaans–San–English*. First edition/Eerste uitgawe. 2004, xi + 377 pp. ISBN 0 7978 2704 8. Johannesburg: Macmillan South Africa Publishers (Pty) Ltd. Prys/Price: R198. (Review in this issue.)
- Herbert Ernst Wiegand. *Internationale Bibliographie 01 zur germanistischen Lexikographie und Wörterbuchforschung. A–H. Mit Berücksichtigung anglistischer, nordistischer, romanistischer, slavistischer und weiterer metalexikographischer Forschungen*. ISBN-13: 978-3-11-013758-3. ISBN-10: 3-11-013758-5. Berlin: Walter de Gruyter. Price: €198. (See review in this issue.)
- Herbert Ernst Wiegand. *Internationale Bibliographie 02 zur germanistischen Lexikographie und Wörterbuchforschung. I–R. Mit Berücksichtigung anglistischer, nordistischer, romanistischer, slavistischer und weiterer metalexikographischer Forschungen*. ISBN-13: 978-3-11-019026-7. ISBN-10: 3-11-019026-5. Berlin: Walter de Gruyter. Price: €198. (See review in this issue.)
- Ladislav Zgusta. *Lexicography Then and Now: Selected Essays*. Edited by Fredric S.F. Dolezal and Thomas B.I. Creamer. 2006, xiv + 404 pp. ISBN-13: 978-3-484-39129-1. ISBN-10: 3-484-39129-4. Lexicographica. Series Maior 129. Tübingen: Max Niemeyer. Price: €118.

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 nuusvystellings 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 type-written A4 pages with double spacing and ample margins (about 2,5 cms). Manuscript should preferably be in electronic form on a (360 KB to 1.44 MB) floppy disk as either ASCII text, fully-formatted Microsoft Word (DOS or Windows) or WordPerfect (DOS or Windows) files. A computer printout of the article should accompany the disk. Each article must be accompanied by an English abstract of 150 to 250 words, and between 10 and 30 English keywords.

2. Graphics

One set of clear original drawings, tables, graphs, diagrams or quality prints thereof must be submitted. The locations of graphics must be clearly indicated in the text.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

HINWEISE UND RICHTLINIEN FÜR AUTOREN

(Nehmen Sie bitte 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*.