

<http://lexikos.journals.ac.za>

Lexikos 24

<http://lexikos.journals.ac.za>

Lexikos 24

Redakteur

Editor

Rufus H. Gouws

Resensieredakteur

Review Editor

T. Harteveld



African Association for Lexicography

AFRILEX-REEKS 24:2014

AFRILEX SERIES 24:2014



BURO VAN DIE WAT

STELLENBOSCH

<http://lexikos.journals.ac.za>

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

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

Uitgewer Publisher

BURO VAN DIE WAT
Posbus 245
7599 STELLENBOSCH

Kopiereg © 2014 deur die uitgewer
Alle regte streng voorbehou
Eerste uitgawe 2014

Tipografie en uitleg deur Tanja Harteveld en Hermien van der Westhuizen
Bandontwerp deur Piet Grobler
Geset in 10 op 12 pt Palatino

Gedruk en gebind deur SUN MeDIA Stellenbosch
Office 5, The Woodmill, Vredenburg Road, Stellenbosch

ISBN 978-0-9814434-9-2
ISSN 1684-4904

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

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

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

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

Lexikos is elektronies beskikbaar by <http://lexikos.journals.ac.za/>

Lexikos is available online at <http://lexikos.journals.ac.za/>

Lexikos is elektronies beskikbaar by Sabinet, AJOL, Ebsco en Proquest

Lexikos is available online from Sabinet, AJOL, Ebsco and Proquest

Indekse Indexes

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

Adviesraad / Advisory Board

Prof. A. Adamska-Salaciak (Pole/Poland)
Prof. H. Béjoint (Frankryk/France)
Prof. H. Chimhundu (Zimbabwe)
Prof. F. Dolezal (VSA/USA)
Prof. R.H. Gouws (RSA)
Prof. R.R.K. Hartmann (Groot-Brittanje/Great Britain)
Prof. M.H. Heliel (Egipte/Egypt)
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)
Dr. J. Tent (Australië/Australia)
Prof. J. Van Keymeulen (België/Belgium)
Prof. P.G.J. van Sterkenburg (Nederland/The Netherlands)
Prof. L.S. Vikør (Noorweë/Norway)
Prof. H.E. Wiegand (Duitsland/Germany)

Redaksiekomitee / Editorial Committee

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

Inhoud / Contents

Voorwoord	x
Foreword	xi
<i>Rufus H. Gouws</i>	
'n Woord van AFRILEX	xii
A Few Words from AFRILEX	xiv
<i>V.M. Mojela</i>	
Redaksionele doelstellings	xvi
Editorial Objectives	xvii

Artikels / Articles

Terminology Development at Tertiary Institutions: A South African Perspective	1
<i>Mariëtta Alberts</i>	
There is No Need for the Terms Polysemy and Homonymy in Lexicography	27
<i>Henning Bergenholtz and Heidi Agerbo</i>	
Explaining Dysfunctional Effects of Lexicographical Communication	36
<i>Herman L. Beyer</i>	
A Functional Analysis of the e-WAT with Specific Focus on the Mobile Version: Towards a Model for Improvement	75
<i>André du Plessis</i>	
A General Lexicographic Model for a Typological Variety of Dictionaries in African Languages	94
<i>Gertrud Faaß, Sonja E. Bosch and Rufus H. Gouws</i>	
Buitetekste in 'n elektroniese gebaretaalwoordeboek	116
<i>Hanelle Fourie Blair</i>	

Article Structures: Moving from Printed to e-Dictionaries <i>Rufus H. Gouws</i>	155
Towards a Southern African English Defining Vocabulary <i>Lorna Hiles</i>	178
The Lemmatization of Loan Words in the <i>isiNdebele-English isiHlathululi-imagama/Dictionary</i> and Their Successful Incorporation into the Language <i>Sponono Mahlangu</i>	186
Example Sentences in Bilingual Specialised Dictionaries Assisting Communication in a Foreign Language <i>Sandro Nielsen</i>	198
The Lemmatisation of Nouns in Tshivenḁa Dictionaries <i>Mashudu Nthambeleni and Nelson Mbulaheni Musehane</i>	214
Elicitation and Arrangement of Conceptual Meanings in the Lexicography of Less Documented Languages <i>Celestino Oriikiriza</i>	225
A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatisation — The <i>Groot Noord-Sotho Woordeboek</i> as a Case in Point <i>D.J. Prinsloo</i>	251
Lexicographic Treatment of Kinship Terms in an English/ Sepedi–Setswana–Sesotho Dictionary with an Amalgamated Lemmalist <i>D.J. Prinsloo</i>	272
The Inclusion of Word Formation in OALD8: The Case of Undefined Run-ons <i>Alenka Vrbinc and Marjeta Vrbinc</i>	291
Leksikovaria / Leksicovaria	
La "main invisible" dans les langues et la confection de dictionnaires locaux <i>Edgard Maillard Ella</i>	310

Projekte / Projects

- Creating a German–Basque Electronic Dictionary for German Learners 331
David Lindemann
- Developing a Campus Slang Dictionary for the University of Botswana 350
Thapelo J. Otlogetswe and Goabilwe Ramaeba

Resensieartikels / Review Articles

- Key Issues in Fuertes–Olivera and Tarp's *Theory and Practice of Specialised Online Dictionaries* 362
Valeria Caruso
- Evalueringstriteria en die interaksie tussen die leksikografie-teorie en -praktyk; die ontwerp van die *Woordeboek van die Afrikaanse Taal* as gevallestudie 378
P.H. Swanepoel

Resensies / Reviews

- Julie Coleman.** *A History of Cant and Slang Dictionaries. Volume I: 1567–1784.* 402
Julie Coleman. *A History of Cant and Slang Dictionaries. Volume II: 1785–1858.*
Julie Coleman. *A History of Cant and Slang Dictionaries. Volume III: 1859–1936.*
Julie Coleman. *A History of Cant and Slang Dictionaries. Volume IV: 1937–1984.*
J.C.M.D. du Plessis
- G.-M. de Schryver et al. (Eds.).** *Oxford Bilingual School Dictionary: IsiXhosa and English* 417
Dion Nkomo
- Mariusz Kamiński.** *A History of the Chambers Dictionary* 423
J.C.M.D. du Plessis

Taasisi ya Taaluma za Kiswahili (TATAKI). <i>Kamusi ya Kiswahili Sanifu</i> <i>Stephen Oluoch</i>	428
Sylwia Wojciechowska. <i>Conceptual Metonymy and Lexicographic Representation</i> <i>José Antonio Jódar Sánchez</i>	431
Publikasieaankondigings / Publication Announcements	442
Voorskrifte aan Skrywers	443
Instructions to Authors	444

Voorwoord

Na die uittrede van dr. Johan du Plessis as redakteur van *Lexikos* na afloop van die publikasie van *Lexikos* 20 in 2010 is daar besluit op die instelling van 'n redaksie van drie lede, te wete proff. Elsabé Taljard, Danie Prinsloo en Rufus Gouws, wat elkeen om die beurt as redakteur van 'n bepaalde uitgawe van *Lexikos* sou optree. In ooreenstemming hiermee was proff. Elsabé Taljard en Danie Prinsloo die redakteurs van onderskeidelik *Lexikos* 21 en 22. In 2013 kon ek weens ander verpligtinge nie die taak oorneem nie en dr. Johan du Plessis het goedgunstiglik ingewillig om as redakteur van *Lexikos* 23 op te tree. Ons hartlike dank aan hom dat hy ondanks sy redakteursaftrede bereid was om weer as redakteur op te tree. Hy het hierdie taak soos altyd in die verlede met groot sukses afgehandel. Vir *Lexikos* 24 word die beplande rotasie van redakteurs voortgesit en word die eerste siklus van die stelsel van drie redakteurs voltooi.

Die drie redakteurs het besluit om in die toekoms telkens gesamentlik verantwoordelik te wees vir elke uitgawe van *Lexikos*. Een van die redakteurs sal as eindredakteur optree, maar sal deur die ander twee redakteurs bygestaan word. Vir *Lexikos* 25 in 2015 het prof. Elsabé Taljard ingewillig om eindredakteur te wees. Alle bydraes vir *Lexikos* word steeds aan die Buro van die Woordboek van die Afrikaanse Taal gestuur by lexikos@sun.ac.za.

In hierdie uitgawe van *Lexikos* is daar weer 'n verskeidenheid bydraes uit verskillende lande en oor verskillende leksikografiese kwessies van verskillende tale wat nogmaals blyke gee van die dinamika van die teoretiese en die praktiese leksikografie. My opregte dank aan al die outeurs vir hulle bydraes — ook daardie outeurs wie se bydraes te laat was vir publikasie in vanjaar se uitgawe van *Lexikos* en wat tot volgende jaar se uitgawe moet oorstaan. Ons skrywers verseker die voortbestaan van *Lexikos* as 'n gesaghebbende internasionale wetenskaplike tydskrif. Bydraes is weer eens aan keuring onderwerp en die belang van die ewekniekeurders se rol sal nooit onderskat word nie. Ook aan hulle my opregte dank.

In die Buro van die WAT was me. Tanja Harteveld die skakelpersoon wat terselfdertyd resensieredakteur was en ook vir die voorbereiding en persklaarmaking van die manuskrip verantwoordelik was. In die set van die artikels is sy bygestaan deur mev. Hermien van der Westhuizen. Hulle kundige en toegewyde hulp het die taak van die redakteur aansienlik makliker gemaak. Baie dankie aan hulle vir die groot taak wat hulle met entoesiasme uitgevoer het.

Mag hierdie uitgawe van *Lexikos* die leksikografie ter wille wees en 'n aktiewe bydrae lewer tot die internasionale leksikografiese gesprek.

Die redakteurs ontvang graag kommentaar, wenke en kritiek van lesers. Dit kan gestuur word aan rhg@sun.ac.za.

Rufus H. Gouws
Redakteur

Foreword

After Dr Johan du Plessis retired as editor of *Lexikos* following the publication of *Lexikos* 20 in 2010, the decision was taken to introduce a system of three editors, i.e. Proff. Elsabé Taljard, Danie Prinsloo and Rufus Gouws. Each one would get a turn to be editor of a given issue of *Lexikos*. In accordance with this decision Proff. Elsabé Taljard and Danie Prinsloo were the editors of *Lexikos* 21 and 22 respectively. Due to other obligations I could not take the responsibility of being editor in 2013 and Dr Johan du Plessis kindly agreed to edit *Lexikos* 23. Our sincere thanks to him for his willingness to take this responsibility in spite of already having retired from this position. As always he performed his duties in an exemplary and successful way. With *Lexikos* 24 the envisaged rotation of editors is continued and the first cycle of the new system is completed.

The three editors have decided to take joint responsibility for each future issue of *Lexikos*. One member of the team will annually be appointed as final editor but that person will be supported by the other team members. Prof. Elsabé Taljard has agreed to be the final editor for *Lexikos* 25 in 2015. All contributions should still be mailed to the Bureau of the Woordboek van die Afrikaanse Taal at lexikos@sun.ac.za.

This issue of *Lexikos* contains a variety of contributions from different countries, dealing with different lexicographic issues of different languages. These contributions yet again illustrate the dynamic nature of both theoretical and practical lexicography. My sincere thanks to all our authors — also those authors whose contributions were too late for publication in this year's volume and have to be kept for next year's issue of *Lexikos*. Our authors ensure the continued success of *Lexikos* as an authoritative international scientific journal. Contributions are subjected to peer reviewing and the significant role of the peer reviewers will never be underestimated. They also deserve our sincere gratitude.

In the office of the Bureau of the WAT Ms. Tanja Hartevelde has been the contact person as well as the review editor and the person responsible for preparing the manuscript for publication. Mrs. Hermien van der Westhuizen assisted her in the typesetting of the articles. Their expert and dedicated assistance made the task of the editor so much easier. Thank you very much to both these ladies for the enthusiastic way in which they performed this big task.

May this issue of *Lexikos* benefit lexicography and make an active contribution to the international lexicographic discussion.

The editors welcome your comments, hints and criticism. Please mail it to rhg@sun.ac.za.

Rufus H. Gouws
Editor

'n Woord van AFRILEX

AFRILEX bedank graag prof. Rufus Gouws, een van die vier erelede van AFRILEX, wat die verantwoordelikheid aanvaar het om redakteur van vanjaar se uitgawe, naamlik *Lexikos* 24, te wees. Prof. Gouws is 'n bekende deskundige op die gebied van taalkunde en leksikografie, 'n skrywer, navorser en akademikus met 'n hoë profiel. Ons weet dat *Lexikos* by hom in goeie hande is. Namens AFRILEX wil ons hom hartlik bedank en ons wens is dat die Almagtige hom lei en krag gee om hierdie groot werk met gemak uit te voer. Ons wil by voorbaat ook reeds proff. Danie Prinsloo en Elsabé Taljard bedank wat die taak by prof. Gouws sal oorneem en in 2015 sal voortgaan met die redakteurskap van hierdie tydskrif. Sowel prof. Taljard as prof. Prinsloo is nie nuut in hierdie taak nie omdat hulle reeds voorheen skitterende werk gelewer het met die redakteurskap van *Lexikos* 21 en 22 in 2011 en 2012 onderskeidelik, nadat dr. Du Plessis in 2010 uitgetree het. Dit beteken dat alhoewel ons ons veteraan *Lexikos*-redakteur, dr. Johan du Plessis, baie sal mis, *Lexikos* sal voortgaan om te groei omdat dit steeds in goeie hande sal wees. Die toekoms van hierdie internasionale tydskrif is baie blink en dit sal in die toekoms steeds voorsien in die ontwikkeling van leksikografie op die vasteland van Afrika.

As mondstuk van die leksikografie het *Lexikos* sedert die eerste uitgawe 'n puik forum gebied vir leksikograwe, taalkundiges en AFRILEX-lede om hulle leksikografiese navorsingsuitsette met nasionale en internasionale kollegas te deel. Hierdie tydskrif het baie bygedra tot die ontwikkeling van leksikografie op die vasteland van Afrika. Die artikels in hierdie jaar se uitgawe van *Lexikos* handhaaf die hoë standaard. In AFRILEX is ons baie bly om te sien hoe die tydskrif groei en met elke nuwe deel van krag tot krag gaan — wat sowel kwantiteit as kwaliteit betref — en dat die tydskrif goed meeding met verskeie ander tydskrifte wat elders in die wêreld gepubliseer word. Hierdie sukses is volledig toe te skryf aan die werk van die redakteurs en die ewekniebeoordelaars wat slapelose nagte deurbring om te verseker dat die bydraes geskik is vir opname in *Lexikos*. Ons wil die skrywers van hierdie bydraes bedank, veral die AFRILEX-lede, wat die leksikografievaandel laat wapper deur *Lexikos* gereeld van gehalte-artikels te voorsien wat die status van die tydskrif jaar na jaar verhoog en in stand hou.

Weer eens wil ons al die AFRILEX-lede wat referate by ons kongresse lewer, aanmoedig om hulle bydraes vir publikasie in hierdie geakkrediteerde tydskrif voor te lê en om hulle navorsingskundigheid met ander kollegas te deel, waaronder die talle wat nie die kongresse kan bywoon nie. *Lexikos* se beskikbaarheid aanlyn het die tydskrif geglobaliseer en dit bemark die tydskrif en die skrywers van die artikels nasionaal en internasionaal en gee aan hulle

blootstelling op die internasionale forums van leksikografienavorsing en kundigheid.

As president van AFRILEX en personeellid van 'n woordeboekeenhede wil ek ook 'n beroep doen op ons leksikografiekollegas in al die Nasionale Woordeboekeenhede in Suid-Afrika om *Lexikos* te sien as 'n platform wat spesifiek daar is vir die publikasie van alle leksikografiese ontwikkeling, uitdagings, ontdekkings en suksesse van die woordeboekeenhede. Dit is 'n geleentheid wat nie misbruik moet word nie.

Maropeng Victor Mojela
President: AFRILEX

A Few Words from AFRILEX

AFRILEX wants to thank Prof Rufus Gouws, one of our four Honorary AFRILEX members, for accepting the responsibility as editor of this year's volume, i.e. *Lexikos* 24. Prof Gouws is a well-known expert in the field of linguistics and lexicography, an author, a researcher and an academic of high profile. In him we all know that *Lexikos* is in good hands. On behalf of AFRILEX, we want to thank him heartily and wish the Almighty guide and give him strength to carry out this heavy job with ease. We also want to thank in advance, Prof DJ Prinsloo and Prof Elsabe Taljard, who will be taking over from Prof Gouws in 2015 to continue the work of editing this journal. Both Prof Taljard and Prof Prinsloo are not new to this task because they previously did excellent work in editing *Lexikos* 21 and 22 in 2011 and 2012 respectively after Dr Du Plessis's retirement in 2010. This simply means that, even though we will miss our veteran *Lexikos* editor, Dr Johan Du Plessis very much, *Lexikos* will continue to grow because it is still in very good hands. The future of this international journal is very bright and it will cater for the development of lexicography on the African continent.

As the mouthpiece of lexicography, *Lexikos* has since its inception created an excellent forum for lexicographers, linguistic scholars and AFRILEX members to share their lexicography research outputs with colleagues, both nationally and internationally. This journal has contributed much to the development of lexicography on the African continent. This year's volume of *Lexikos* is still keeping the standard high with regards to the quality of the published articles. In AFRILEX we are very pleased to see this journal growing from strength to strength in both quantity and quality with the advent of every volume, competing effectively with various journals published elsewhere on the globe. All these successes are solely ascribed to the work of its editors and the peer reviewers who spend sleepless nights making sure that the *Lexikos* manuscripts are at the level suitable for inclusion in this journal. We also want to thank the authors of the articles, especially the AFRILEX members, who keep the lexicography flag high by continuously supplying *Lexikos* with quality articles to elevate and sustain the status of this journal year by year.

Once-more we want to encourage all AFRILEX members who read papers at AFRILEX conferences to prepare their papers for publication in this accredited journal and to share their research expertise with other scholars, including the overwhelming majority who could not attend the conference. The availability of *Lexikos* online has globalized this journal, and this is marketing the journal as well as the authors of its articles nationally and internationally to expose them to the international forums of lexicography research and scholarship.

As AFRILEX President and member of staff of a Lexicography Unit I once

again call upon our lexicography colleagues in all the National Lexicography Units in South Africa to view *Lexikos* as a platform primarily established to publish all lexicography developments, challenges, discoveries and successes pertaining to all dictionary units, and this is the opportunity not to be misused.

Maropeng Victor Mojela
President: AFRILEX

Redaksionele doelstellings

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

Die strewe van die AFRILEX-reeks is:

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

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

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

<http://lexikos.journals.ac.za>

Terminology Development at Tertiary Institutions: A South African Perspective

Mariëtta Alberts, *Research Unit for Languages and Literature in SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa (AlbertsMarietta@gmail.com)*

Abstract: There is a dire need in South Africa for multilingual polythematic terminology. Currently no tertiary institution presents terminology theory and practice as a fully-fledged subject and there is also no sufficient mechanism for the training of terminologists. This situation provides tertiary institutions the opportunity to position themselves in terms of terminology training and development. Terminology centres at tertiary institutions could be of great value for language as well as all other subject-related departments. The language departments, where such terminology centres could possibly be housed, already have facilities to teach some or all of the official languages of South Africa at tertiary level. Language students could receive postgraduate (and even undergraduate) training in the theory and practice of terminology. For practical work, they could collect terminology at subject-related departments and take it to the terminology centres, where source language terms could be terminographically treated. Subject specialists (i.e. lecturers teaching different subjects) would have to be consulted when defining the concepts and would have to assist language students and their lecturers to supply term equivalents in target languages. This process would enable subject-related and language departments to cooperate in different domains. Language and language departments would therefore become useful to different subject areas. Multilingual polythematic terms with definitions could be available to students who are not proficient in English or Afrikaans. They would therefore have access to study material in their first language. The terminology could be disseminated internally and (inter)nationally through the Internet, which would make it accessible to any other student. This process could best be managed and executed by dedicated terminology centres.

Keywords: LEGISLATION, MULTILINGUAL TERMINOLOGY, POLYTHEMATIC TERMINOLOGY, SOURCE LANGUAGE, STANDARDISATION, TARGET LANGUAGE, TERMINOGRAPHY, TERMINOLOGY, TERMINOLOGY CENTRE, TERMINOLOGY DEVELOPMENT, TERMINOLOGY TRAINING

Opsomming: Terminologieontwikkeling aan tersiêre instansies: 'n Suid-Afrikaanse perspektief. Daar bestaan in Suid-Afrika 'n groot behoefte aan meertalige terminologie op verskeie vakgebiede. Tans bied geen tersiêre instansie terminologieteorie en -praktyk as volwaardige vak aan nie en daar is nie 'n doeltreffende meganisme waarvolgens terminoloë opgelei kan word nie. Hierdie situasie skep die geleentheid vir tersiêre instansies om hulself te posisioneer ten opsigte van terminologieopleiding en -ontwikkeling. Terminologiesentra aan tersiêre instansies sal vir taal- sowel as alle ander vakgerigte departemente van onskatbare waarde wees. Die

taaldepartemente, waar sodanige sentra waarskynlik gehuisves sou kon word, beskik reeds oor fasiliteite om sommige of alle amptelike tale van Suid-Afrika op tersiêre vlak aan te bied. Taalstude-
dente kan nagraadse (en selfs voorgraadse) opleiding kry in terminologieteorie en -praktyk. As prak-
tiese werk sou hulle terminologie by vakgerigte departemente kon gaan versamel en na die sentra
neem waar die brontaalterme terminografies hanteer kan word. Vakkundiges (bv. lektore wat vakke
aanbied) sal met die definiëring van konsepte behulpsaam moet wees en sal saam met die taalstu-
dente en -dosente vertaalekwivalente in doeltale moet verskaf. Dié proses sal meebring dat vakkun-
diges en taalkundiges oor vakgrense heen saamwerk. Dit sal daartoe lei dat taal en taaldepartemente
vir ander vakterreine van waarde sal raak. Meertalige vakgerigte terme met definisies sal beskik-
baar wees vir studente wat nie Engels of Afrikaans magtig is nie. Hulle sal dus toegang tot studie-
materiaal in hul eerste taal kan kry. Die terminologie kan intern deur die tersiêre instansies en
(inter)nasionaal deur die Internet versprei word, wat dit toeganklik sal maak vir enige ander stu-
dent. Hierdie proses kan ten beste deur doelgerigte terminologiesentra beheer en uitgevoer word.

Sleutelwoorde: BRONTAAL, DOELTAAL, MEERTALIGE TERMINOLOGIE, STANDAARDI-
SERING, TERMINOGRAFIE, TERMINOLOGIE, TERMINOLOGIEOPLEIDING, TERMINOLOGIE-
ONTWIKKELING, TERMINOLOGIESENTRUM, VAKGERIGTE TERMINOLOGIE, WETGEWING

1. Introduction

Terminology plays a pivotal role in the teaching and learning of subject matter. Terminology is a tool for communication in languages for special purposes, e.g. subject areas (mathematics, physics, chemistry, botany, zoology) and domains (art, music, sport).

Specialised information is encoded in a source language to be conveyed (transmitted) from a sender as communication source (e.g. lecturer) to a receiver as recipient of information (e.g. student), who will decode the message and then react on the stimulus received. Terminology allows subject-related communication between subject specialists and between subject specialists and lay-people (cf. Alberts 2012b).

Lecturers (as subject specialists) teach a specific subject (e.g. physics) to students (laypeople) to equip them to become subject specialists themselves and to enable them to serve as knowledgeable workers in a vocation-specific environment. This argument is only valid if the subject specialist is able to convey information to the student. Lecturers should be educated to respond to "desired" mother-tongue education (cf. Maseko 2012: 5).

Most South African tertiary institutions primarily use English as language of learning and teaching, and sometimes Afrikaans. It is argued that Afrikaans had the opportunity to develop as an academic language during the previous political dispensation. Even African languages as subjects are taught largely through the medium of other languages (cf. Maseko 2012: 3).

There are several arguments underpinning the value of instruction through the mother tongue or first language of both trainer and trainee. It is, however, currently not possible to supply mother-tongue/first language education in

South Africa during the tertiary phase of education. There are several reasons for this situation, but this article will only focus on ways to solve the need for multilingual terminology in various subject-related areas at tertiary institutions.

2. Relevant legislation

According to Census 2011 results, only 9,6% of the South African population speak English as their first language, while the rest of the population speak other languages as their first language. Given that the majority of South Africans are not first-language speakers of English, the implication is that they would have limited or no access to information, an important commodity for participation and democracy (cf. Statistics South Africa 2013: iii). The value of available standardised multilingual polythematic terminology cannot be over-emphasised. The need for such terminology is probably greater in South Africa than in other countries, seeing that we have a multilingual dispensation encapsulated in the Constitution of the Republic of South Africa, 1996.

The Constitution of the Republic of South Africa, 1996, provides for the equitable use of all the official languages of the country, and for the promotion and development of the historically marginalised indigenous languages. The National Language Policy Framework of March 2003 gives effect to the constitutional rights regarding language usage and development (cf. Department of Sport, Recreation, Arts and Culture 2005: 1, 3).

Section 6(2) of the Constitution of the Republic of South Africa, 1996, recognises the historically diminished use and status of the indigenous languages of South Africa, and compels the state to take practical steps to design mechanisms to elevate the status and to advance the use of these languages (cf. Department of Sport, Recreation, Arts and Culture 2005: 3). The Constitution further requires all official languages to enjoy parity of esteem and to be treated equitably. It also provides for national government to regulate and monitor the use of official languages by legislative and other means (cf. Department of Arts and Culture 2011: 7; *Government Gazette* 2012: 2, 4). The Use of Official Languages Act, 2012 (Act No. 12 of 2012) stipulates that every national department, national public entity and national public enterprise should promote the use of official languages in facilitating access to services and information. They should furthermore develop a language policy regarding their use of official languages, amongst others, in effectively communicating with the public, official notices, government publications and inter- and intra-government communications (cf. *Government Gazette* 2012: 6; cf. Regulation 7(1)(d) *Government Gazette* 2013: 25). The development of multilingual terminology for the official languages gives effect to the constitutional rights of languages regarding their usage and their further development.

It is clear that political change in 1994 provided a legislative context for transformation in all phases of education and a redress of imbalances resultant from the previous dispensation (cf. Maseko 2012: 3). Language, amongst other

issues, is seen as critical in fostering transformation in higher education and in society. The role of English and Afrikaans as languages of higher learning is acknowledged but tertiary institutions should not exclude nor discriminate against speakers of other languages (cf. Maseko 2012: 4). The development of all official languages to address imbalances of the past is a legislative imperative and is reflected in the legislative context of the Constitution of the Republic of South Africa (1996); Higher Education Act (1997); National Plan for Higher Education (2001); Language Policy for Higher Education (2002); the Ndebele Report on the Development of Indigenous African Languages as Mediums of Instruction in Higher Education (2003); Development of Indigenous languages as Mediums of Instruction in Higher Education (Report compiled by the Ministerial Committee appointed by the Ministry of Education) (2003); Humanities and Social Sciences Charter (2011) and the Green Paper on Post-secondary School Education and Training (2012) (cf. Maseko 2012: 4-5, 34-36).

Section 27(2) of the Higher Education Act (101 of 1997) requires the Minister to determine language policy for Higher Education. Subject to the policy, Councils of public higher education institutions, with the concurrence of their Senates, must determine the language policy of an institution and publish and make the policy available on request. The main objective of the Language Policy for Higher Education (Section 6, 2002) is to promote multilingualism in institutional policies and practices of universities:

The challenge facing higher education is to ensure the simultaneous development of a multilingual environment in which all our languages are developed as academic/scientific languages, while at the same time ensuring that the existing languages of instruction do not serve as a barrier to access and success (cf. Department of Higher Education and Training 2010: 3-6).

The Green Paper on Post-secondary school Education and Training, 2012, par. 6.10 recognises the unfortunate position of African languages in South African Higher Education and provides for African languages to be utilised across disciplines at tertiary institutions (cf. Department of Higher Education and Training 2012; Maseko 2012: 6). No subject could, however, be taught without relevant subject-specific terminology. Without proper subject-specific terminology in the language(s) of choice, the principle of having a shared language of learning and teaching would remain a pipedream.

Terminology development adheres to the language policy of the country. In the previous dispensation, with a bilingual policy, terms were supplied in English and Afrikaans. Since 1994 the Republic of South Africa has had a multilingual policy and therefore terminology should be available in the 11 official languages. Terms should also be available in South African Sign Language (SASL) — in this case signs to denote the underlying concepts (cf. Alberts 2012a).

The following arguments are valid for advancing the use of multilingual terminology in higher education:

— Pedagogic argument — the link between language and cognition in terms

of multilingualism in education.

- Economic argument — economic participation both locally and globally, and the resultant social mobility.
- Socio-cultural argument — identity of "self" in the context of "others".
- Linguistic argument — language development/study which, if not undertaken, is a threat to linguistic diversity, and could lead to language decline or death.
- Political argument — nation building, which emphasises unity in diversity; equity, access, success, redress, language rights and transformation (cf. Maseko 2012: 10).

The Language Policy for Higher Education (2002) highlights the need to promote South African languages for use in instruction in higher education. It also stresses the need to develop strategies for promoting proficiency in languages of tuition. It furthermore encourages all tertiary institutions to consider ways of promoting multilingualism and requires institutions to indicate in their plans what strategies have been put in place (cf. Department of Higher Education and Training 2010: 7-8).

A 2011–2012 survey on language choices regarding the language policies of tertiary institutions found that 17 out of 23 universities have published their language policies and all institutions surveyed have selected at least one indigenous African language as another language of learning and training. The implementation plans, especially on the promotion and development of the African languages were, however, lacking (cf. Maseko 2012: 8).

Afrikaans is an example of a language that was initially regarded as a "kitchen language" but which developed into a functional language in *inter alia* science, technology, economy and education. There is therefore no reason why the other official languages could not develop into languages of higher learning. Because of the hegemony of English, the other official languages often struggle with terminology development. Most terms denoting new concepts reach South Africa through the medium of English and special efforts need to be made to translate these terms into the various official languages (cf. Alberts 2012a).

By developing subject-specific terminology in all official languages and by compiling multilingual polythematic term lists/glossaries/dictionaries and other supporting teaching resources (i.e. multilingual study material that incorporates the multilingual terminology) the process to develop the official languages into scientific languages will be facilitated. Academic proficiency in the language of learning and teaching will also be promoted.

3. Terminological principles

It is a given that one understands concepts best in one's first language. It is also

easier to convey information (e.g. teach/train) through the medium of the first language of both the lecturer/trainer and the student/trainee. Lecturers who understand the subject matter in their first language will be able to convey the information to students. Conceptualisation is also paramount in the first language/mother tongue (cf. Alberts 2012a).

Students at tertiary institutions are confronted by different subjects that contain difficult subject matter to be comprehended. What makes this issue even worse is that these subjects are often taught in many lecturers' and students' second or third language. Terminology is not always available in the different subject areas and the different official languages. The same concept is often denoted by means of different terms (cf. digestive tracts/gastrointestinal tract/alimentary canal), which leads to confusion. The increasing use of non-standardised terms in the lecture hall often has dire consequences for students and for the subject and language in question.

When terminology is developed, one should adhere to basic terminological principles:

- A concept is denoted by means of a term.
- A term has an exact meaning and has to denote the features of the concept it represents.
- A term therefore refers to a definite concept which is clearly defined within specific parameters.
- A term is a linguistic representation of a mental construct (cf. Alberts 2012b).

Terminology can be regarded as any of the following:

- a collection of terms in a specialised subject area or domain
- the methods used to harvest, document, systematise, standardise and disseminate terms
- the science researching the nature of terms, their form and communicative function (Cluver 1983: 7).

The primary aim of terminology is to promote communication in scientific and technological environments. Terms are documented and systematised in order to compile technical dictionaries, language for special purpose (LSP) dictionaries, glossaries, terminology lists, electronic term banks, or online dictionaries (Alberts 2012b).

Terminography is an interdisciplinary practice that combines areas such as lexicographical principles, linguistics, information science, ontologies, computational linguistics, computer science, etc.

Terminology uses the grammar and orthography (cf. *tree* (en) *vs.* *Baum* (de) (nouns are capitalised in German)) of the standard variety of the language.

Terms are created according to the word-forming principles of the standard variety of the language (cf. Alberts 2012b).

Specialised terms are needed

- that provide access to texts in other languages, e.g. source language (SL) to target language(s) (TL(s))
- that have exact meaning with no emotional connotations
- that function within a specific conceptual cluster
- that are standardised (e.g. 1 = one, een, uno, eins, kunye, nye, nngwe, ...)
- that always denote the same concept (e.g.
 - subtract (-)** *v.* aftrek; ntsha; ntsha; tlosa; thabatha; susa; khokha; khupha; u tusa; susa)
- that are harmonised in related (inter)national languages (e.g.
 - macroinstruction** (en); makroinstruksie (af); makroinstruktion (da); macroinstrução (pt))
- that are harmonised in related South African languages (e.g.
 - count** (*v.*): **tel** (af); **bala** (ns); **bala** (ss); **bala** (tn); **bala** (xh); **bala** (zu); **bala** (nd); **-vhala** (ve); ku hlayela (ts)
 - litre** (*n.*) <ℓ>: **liter** (af); **ilitha** (xh); **ilitha** (zu); **ilitha** (sw); **ilitha** (nd); **litha** (ve); **litara** (ns); **litara** (ss); **litara** (tn); **litara** (ts).

Concepts should be defined in the source language (SL) and terms should be created in the SL to denote the concept (process: primary term creation). Only then term equivalents could be supplied in target language(s) (TLs) (process: secondary term creation) (cf. Alberts 2012a), e.g.

straight line (en) *The shortest distance between two points.* reguit lyn (af); umugqa ogonde thwi (zu); umgca ongqalileyo (xh); umugca locondzile (ss); umudathwi (nd); molatlhamalalo (tn); mothalothwii (ns); mola o otlolohileng (st); mutalo tswititi (ve); layini-thwi; layini yo ololoka (ts).

The terminographical process entails the following:

- demarcation of the subject area or domain
- determining of the target user group
- collection of terms from relevant sources (e.g. text books, dissertations, journals, study material, etc.)
- excerption of relevant terms, i.e. harvesting of source language (SL) terms
- defining of terms in the SL (e.g. English) and providing of contextual sentences and related information (i.e. parts of speech)
- verifying of the SL information with subject specialists

- changing of the database according to input received
- adding of target language (TL) equivalent(s) and related information (e.g. definitions, example sentences, parts of speech, etc.)
- verifying of TL term equivalents and related information with the relevant Technical Committee (TC) of the relevant National Language Body (NLB) for its stamp of approval of the linguistic content
- adjusting and updating of the database according to suggestions received
- dissemination of the SL terms and TL equivalents with related data to target users via hard copy publications (term lists, glossaries), electronic dictionaries (CD ROM, online, Internet), documents containing terms (text-books, study guides, dissertations, manuals, journals, technical brochures, etc.), mass communication media (radio, press, television, etc.).

Terminology assists with subject-related communication. Terminology is documented with the aim of providing subject specialists (lecturers) and laypeople (students) with standardised terms that denote the corresponding concepts.

Terminology should be standardised to allow exact communication. The standardisation process entails that the majority of subject specialists should agree to denote a definite concept with a specific term through

- the limiting of meaning by means of a proper definition within the conceptual cluster of the subject area — one term should denote one concept and one should try to avoid the use of synonyms
- the systematisation and usage of language rules, e.g. grammar, word-forming principles, spelling and orthographic rules, and the consultation of linguists, since they play an important role in the standardisation process
- the submission of the SL terms and TL term equivalents to the relevant National Language Body (NLB) for the linguistic verification and approval of the terms
- the dissemination of terms (dictionaries or online facilities) to ensure usage and the penetration of terms in the relevant subject area or domain and in different languages (cf. Cluver 1982: 81; Alberts 2012a).

Terms that are not standardised, give rise to communication problems. It should, however, be noted that standardisation is a process that takes time. Terms will only become standardised when they are used properly and penetrate the subject area or domain and the specific language through frequent usage.

4. Models for terminology development for the tertiary phase of education

One could ask what could be done to solve terminology needs and to develop terminology for usage by lecturers and students at tertiary institutions. Dedi-

cated terminology institutions, publishing houses, private initiatives and tertiary institutions could all develop terminology to address the need for multilingual polythematic terminology at tertiary institutions. Three terminographical models will be discussed, but special emphasis will be placed on terminology training and development by tertiary institutions themselves. An integrated working model will also be presented.

4.1 Dedicated terminology offices

There are several terminology offices in South Africa that are dedicated to the development of vocation-specific terminology. The following offices serve as examples of dedicated terminology offices:

4.1.1 Terminology Coordination Section

Subject-oriented terminography is done by the Terminology Coordination Section (TCS), National Language Service (NLS), Department of Arts and Culture (DAC). The Terminology Coordination Section is the national terminology office, with trained terminologists for each of the official South African languages. The TCS documents multilingual terminology in a variety of subject areas in a centralised national term bank. The TCS registers national projects in a national registry to avoid duplication of projects.

Terminologists usually receive in-house training presented by senior terminologists and external experts in terminology and lexicography. In 1995 the newly appointed African language terminologists, for instance, received intensive training presented by the Head: Systems Development and Research, who at the time was also teaching a terminology and lexicography module as part-time lecturer at the Department of African Languages, University of Pretoria. In 2007, as Director: Terminology Development and Standardisation at PanSALB, she also presented training modules on translation-oriented terminography (TOT) (cf. ISO 2002) to all language practitioners employed by the National Language Services, DAC.

Between 2001 and 2002 three senior terminologists of the TCS, NLS trained collaborators in all nine provinces to assist TCS with terminology development.

4.1.2 Other organisations

Other government, semi-government and non-governmental organisations (NGOs) such as National Parliament, the Department of Justice, Statistics South Africa, GCIS, Eskom, Tshwane Municipality, etc. also compile term lists dedicated to vocation-specific domains, e.g.

- National Parliament, Cape Town — the language practitioners (i.e. interpreters) compile term lists in various domains dealt with by parliament. The

term lists are compiled in all official languages and terminology development is also done for South African Sign Language.

- State Law Adviser's Office, Department of Justice — the language practitioners compile term lists in all official languages. The terminology relates to legislation.
- Statistics South Africa — language practitioners compile term lists relating to statistics. These term lists assist the office when surveys need to be done.
- Provincial Departments and municipalities — language practitioners compile various term lists in the relevant provincial languages.

These and several other national and provincial offices received training on terminological principles and practice between 2002 and 2010, presented by the former Director: Terminology Development and Standardisation, PanSALB. She was also instrumental in the establishment of the Language Unit at the office of the State Law Adviser, Department of Justice.

The following model for the development of terminology represents the subject-oriented terminography process by a dedicated terminology office (cf. Diagram 1):

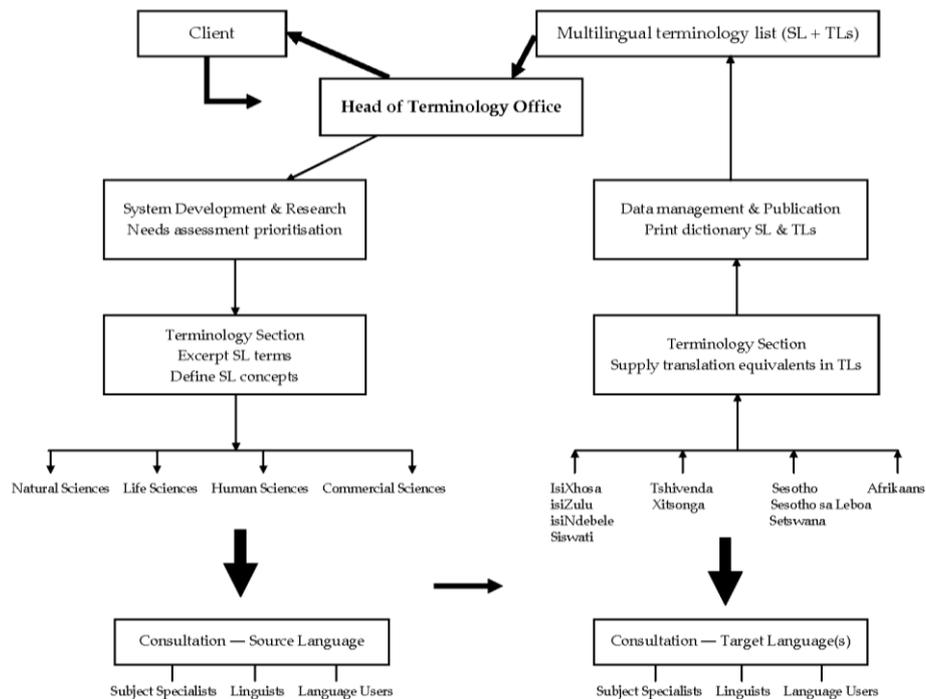


Diagram 1: Subject-oriented terminography process by dedicated terminology office

The terminographical process described should be similar for all offices mentioned above (cf. Diagram 1 and paragraph 3):

- A term list should be compiled in response to a determined need. The need could be expressed by an external body, or the terminology office could determine a need by means of a needs assessment process.
- The subject area is demarcated and the target users determined.
- Source language (SL) terms are collected from relevant resources.
- SL terms are defined in consultation with subject specialists.
- The SL National Language Body verifies and authenticates the linguistic correctness of the SL in terms of word-forming principles, orthography, etc.
- The database is adjusted according to suggestions received.
- Target language (TL) equivalents and definitions are provided in consultation with subject specialists.
- The TL National Language Bodies verify and authenticate the linguistic correctness of the TL term equivalents in terms of grammar, word-forming principles, orthography, etc.
- The database is adjusted and updated according to suggestions received.
- The requested multilingual term list is presented to the external body (e.g. tertiary institution) or published in printed format for general usage by target users, or disseminated online (i.e. Internet).

4.2 Independent terminology centres

There are also private initiatives that provide terminology for tertiary institutions, e.g. projects by the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) (South African Academy for Science and the Arts), the Centre for Political and Related Terminology in Southern African Context (CEPTSA) and the Centre for Legal Terminology in African languages (CLTAL). These centres all collaborate with the Terminology Coordination Section, National Lexicography Units, National Language Bodies and publishing houses.

- SAAWK — A bilingual Arts and Crafts dictionary was recently compiled by experts that worked under the auspices of the SAAWK. The dictionary has been completed and will be published soon. The ideal was to create a multilingual South African Arts Dictionary but the bilingual version was compiled first, since it was established in the work on several multilingual term lists that either the English or Afrikaans subject matter is freely consulted by the African language practitioners in the translation/creation of equivalents in their respective first languages (cf. SAAWK n.d.: 1). As art is universal, so its vocabulary is international and therefore there are numer-

ous loanwords and transliterated concepts in the arts terminology, cf.

art rupestre <see rock art>
art rupestre <kyk rotskuns>

abozzo 1: *The first underpinning (outline or drawing) on a canvas.*

abozzo 1: *Die eerste onderskildering (omlyning of tekening) op 'n doek.*

abozzo 2: *A block of stone, a lump of clay or chunk of wood that has been reduced to a rough form of the ultimate work.*

abozzo 2: *'n Klipblok, kleimassa of houtstomp wat tot 'n ruwe vorm van die uiteindelike werk verklein is*

cartouche 1: *An ornamental panel in the form of a scroll or sheet of paper.*

cartouche 1: *'n Ornamentele paneel in die vorm van 'n rol of vel papier.*

cartouche 2: *An antique Egyptian signature in the form of a picture.*

cartouche 2: *'n Antieke Egiptiese handtekening in prentvorm.*

For the benefit of students, the name of the sub-discipline of art to which an entry belongs, was added, e.g. *Architecture, Drawing, Painting, Ceramics...* (cf. SAAWK n.d.: 5).

- CEPTSA is an independent centre that works under the auspices of the University of Johannesburg. A core group of subject specialists work in collaboration with a terminologist and linguist. The committee published a bilingual English/Afrikaans dictionary as well as a revised version on political terminology (cf. CEPTSA 2002; Botha, Le Clus and Venter 2011).

CEPTSA compiled an English SL term list with 1 000 core terms and explanations and the data was translated into Afrikaans, Zulu and Northern Sotho (cf. Le Clus, Botha, Venter and Reyneke n.d.), cf.

majority rule. *The exercise of power according to the will of the majority.*

meerderheidsheerskappy. *Magsuitoefening volgens die wil van die meerderheid.*

ukubusa ngezwi leningi. *Ukusetshenziswa kwamandla ngokuya ngentando yeningi.*

pušo ya bontši. *Tiragatšo ya maatla go ya ka thato ya bontši.*

A draft list containing English, Afrikaans and four of the African languages (Northern Sotho, Zulu, Xhosa and Tswana) was also compiled, cf.

abolition. *The act of doing away with a practice such as capital punishment or slavery.*

afskaffing. *Die handeling om 'n praktyk soos die doodstraf of slawerny tot niet te maak.*

phedišo. *Modiro wa go fediša mokgwa wa go swana le kotlo ya lehu goba bokgoba.*

ukuchithwa. *Isenzo sokuchitha isejwayezi esithile njengokujeziswa ngakwezomnotho noma ubugqila.*

ukubhangiswa. *Isenzo sokuphelisa inkqubo enjengeyesigwebo sentambo okanye ubukhoboka.*

phediso. *Kgato ya go fedisa/khutlisa tsamaiso/tiragalo jaaka katlholelo leso kgotsa bokgoba.*

The committee recently completed an explanatory list consisting of 2 500

English SL terms. This list has already been supplied with Afrikaans TL term equivalents and explanations. The bilingual explanatory list has been presented to African language terminologists to submit TL equivalents and explanations. The translation of the term equivalents of CEPTSA's new *Verklarende Politieke Woordeboek/Explanatory Dictionary of Politics* is done in collaboration with translators of the Language Units of the University of Johannesburg and the University of Stellenbosch and Juta Publishing House.

- CLTAL is an independent centre that works on criminal law, criminal procedural law and law of evidence. The compilation of the multilingual dictionary is done in cooperation with subject specialists such as lawyers, court interpreters and academics as well as language practitioners, such as linguists, terminologists and translators. CLTAL started with Northern Sotho as a first African language project in combination with English and Afrikaans, e.g.

premeditated murder	polaopeakanywa	voorbegadte moord
polygraph	seutoliamaaka	poligraaf

Several legal professionals and linguists representing other African languages attend the quarterly meetings of the CLTAL and have already started supplying term equivalents and definitions in the other official languages, such as Tswana, South Sotho, Venda, and Tsonga. The first CLTAL publication will be a bilingual English–Afrikaans explanatory dictionary. CLTAL, however, aims to publish multilingual explanatory dictionaries containing all eleven official languages with the aid of its publisher (cf. Alberts 2008).

These centres follow the same terminographical process as described in paragraphs 3 and 4.1. The following model for the development of terminology represents the subject-oriented terminography process followed by independent terminology centres where internal collaborators work closely with external collaborators (cf. Diagram 2):

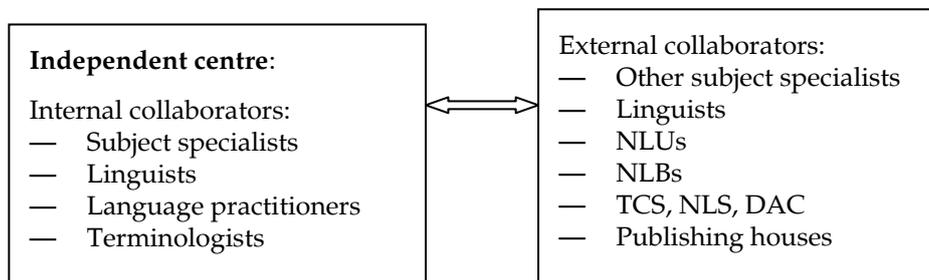


Diagram 2: Subject-oriented terminography process by independent terminology centres

The majority of the work relating to the terminographical process (cf. paragraphs 3 and 4.1 and Diagram 1) is done by the subject specialists themselves and they only receive minor linguistic and terminographical assistance from the terminologist and/or linguist assigned to the projects.

4.3 Terminology centres at tertiary institutions

There are several tertiary institutions that are already realising the value of multilingual polythematic terminology in the under- and postgraduate environment. Yet, a lot of spade work still needs to be done by several major universities before they could really embark on the development of terminology.

4.3.1 Current terminology work at tertiary institutions

The terminographical work done by the Stellenbosch University Language Centre is an excellent example of the kind of terminological assistance that could be provided to lecturers and students.

In February 2006 the former Director: Terminology Development and Standardisation, PanSALB was invited to the Stellenbosch University Language Centre to present a three-day training session on terminological principles, procedures and practices of terminography in African languages (cf. Sibula n.d.: 24).

Since the training, the Unit for IsiXhosa of this Language Centre had already compiled several trilingual Xhosa–English–Afrikaans terminology lists. The terminology lists include three generic term lists, i.e. *Generic Academic list*, *Generic Administrative list* and a *List of Faculties, Departments, Academic Divisions together with Administration and Services* — the typical terminology students would need to orientate themselves when enrolling at the university. Five trilingual subject-specific term lists have also already been compiled, namely *Social Work*, *Sociology*, *Psychology*, *Law* and *Economic and Management Sciences*, cf.

- ❑ **normative** {establishing a norm or standard}
ummiselo-sithethe {kukumisela isithethe nomgangatho}
normatief {vasstelling van 'n norm of standaard}
- ❑ **adolescence** {life phase beginning at puberty and ending in adulthood}
ukufikisa {isigaba sobomi esiqala ekufikeleleni ebudaleni nesiphela ngobuntu obudala}
adolessensie {lebensfase wat begin by puberteit en eindig in volwasenheid}

Three trilingual sport-related term lists were also compiled: *Soccer*, *Cricket* and *Rugby*.

The terminologist consults experts to obtain terminological data. The sub-

ject specialists submit core and problematic concepts and senior students assist with the defining of terms in SL (English). The SL terms and definitions are then provided with Xhosa and Afrikaans term equivalents and definitions. The IsiXhosa Unit is responsible for term creation processes and quality control. The various departments decide how to market and disseminate the terminology lists, e.g. online or hard copies (cf. Sibula n.d.: 4-28).

In 2009 the former Director: Terminology Development and Standardisation, PanSALB was invited to Rhodes University, Grahamstown to present terminology training to enable the University to put multilingual terminological data on various subjects on a blog for utilisation by academics and students.

Rhodes University introduced several programmes to develop subject-related terminology:

- Teaching programme — this programme consists of a generic course in Xhosa as an additional language in subject-related courses such as pharmacy, law, journalism and media studies, education and psychology.
- Terminology lists — English–Xhosa glossaries or term lists are compiled to support learning in English as language of learning and training. Term lists were already compiled for Earth Science, ICT, Politics and Cell Biology.
- Teaching of Xhosa — the teaching of, and in Xhosa as first language in subjects such as linguistics, literary studies, translation studies, lexicography and sociolinguistics.
- Postgraduate research — students are encouraged to do their subject-related research through medium of Xhosa (cf. Maseko 2012: 17).

The former Director: Terminology Development and Standardisation, PanSALB presented either terminology training or urged several other tertiary institutions to start terminology development programmes at their institutions, such as University of Pretoria (2005–2007); PRAESA, University of Cape Town (2005); Tshwane University of Technology (2006); Language Research and Development Centre, Nelson Mandela Metropolitan University (2008); University of South Africa (2009); University of the Free State (2010), and Cape Peninsula University of Technology (2012).

The following lessons were learned from existing terminology-related centres at tertiary institutions:

- Terminology is a strategic resource.
- Terminology development makes a contribution to the development and promotion of the official languages.
- Terminology development is a direct response to the Republic of South Africa's Constitution regarding the development of the official languages.
- All official languages could be developed for use in higher education.

- The development of African languages into languages of learning and training is a difficult and cumbersome but necessary exercise.
- Interactive collaboration is needed during the terminology development process.
- Students use and appreciate the terms and definitions provided in the official African languages.
- Lecturers who are monolingual could promote the use of other languages in class by using bilingual or multilingual teaching aids.
- Language is a symbol of national pride and the use of the indigenous languages builds confidence (cf. Alberts 2012b; Sibula n.d.: 3; Maseko 2012: 32).

4.4 An integrated model: the establishment of terminology centres at tertiary institutions

It is proposed that terminology centres should be developed at language centres or schools of languages at tertiary institutions to present terminology training, to do terminology and terminography-related research and to assist subject-related faculties with the compilation and dissemination of subject-related term lists. The aim should be to train terminologists who could assist with the development of the eleven official languages into functional languages of learning and training and to develop multilingual polythematic terminology.

4.4.1 Terminology training

The latest legislation regarding the use of languages (Act of 2013) ensures job creation possibilities for language practitioners (cf. Government Gazette 2012; 2013). But despite legislation to enhance the language practice, it seems as if fewer students enrol for language courses. The terminology practice urgently needs trained and knowledgeable terminologists.

Terminology training is of importance for language practitioners who would like to work at the Terminology Coordination Section of the National Language Service, Department of Arts and Culture or at any of the language units at national and provincial government departments that need to adhere to current legislation. All government institutions will have to develop the terminology of the subjects related to their fields of expertise and vocation. These terms will have to be in at least three of the official languages.

Only a few tertiary institutions in South Africa are currently dealing with terminology training. Most of the terminology courses are modules within either a linguistics course, a course in translation studies or in lexicography. Given the urgent need for terminology development, there should be dedicated terminology courses stretching from undergraduate level to doctoral studies. These courses should form part of the curricula of schools of languages. Stu-

dents studying any course within a school of languages, such as Afrikaans, African Languages, English, French, German, Linguistics, Roman Languages, etc., should be able to do a theoretical and practical course in terminology and terminography.

Terminology students should get training in *inter alia*:

- the theoretical principles and practice that form the basis of terminology work and terminography
- metalexigraphy
- the difference between lexicography and terminography
- dictionary structures (macro-, micro- and medio-structures)
- typology of dictionaries
- the function of dictionaries
- ontological relationship between concepts
- primary term creation (SL)
- secondary term creation (TL)
- communicative value of terminology
- cultural differences
- compilation of monolingual, bilingual and multilingual dictionaries (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

Terminology students should also get practical experience in terminography work by:

- applying their acquired linguistic knowledge and terminographical skills;
- excerpting subject-related terms from various resources (e.g. text books, study material, journals, dissertations, etc.) — this is an exercise in terminology harvesting methods;
- defining terms to ensure that the content of the concept is reflected in the definition — this is an exercise in the usage of different definition methods;
- doing assignments on the compilation of a source language term list of a specific domain — this is an exercise in primary term creation and any of the languages could be used as source language;
- doing an assignment in collaboration with other students by compiling a bilingual or multilingual term list — this is an exercise in secondary term creation;
- evaluating the compiled term lists of fellow students — this is an exercise in

dictionary criticism and the students can learn from their own mistakes and those of other students (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

The value of terminology training for students:

- They would recognise terminology work and terminography as a career opportunity.
- They would
 - be trained in the basic theories, principles and practice that govern terminology and terminography;
 - be trained in the practice of the compilation of different types of dictionaries;
 - learn the value of collaborative work;
 - get to know the typical working environment of a terminologist and what to expect when such a career is pursued;
 - learn how to deal with peer group criticism in a constructive way;
 - learn the importance of working in collaboration with and consulting subject specialists, linguists, language practitioners, language structures (e.g. NLBs, NLUs, TCS);
 - familiarise themselves with different language structures (TCS, NLBs, NLUs, PLCs) and language units and to differentiate between these structures;
 - be skilled to work at terminology-related institutions and they would require less in-service training.
- Students interested in a career in terminology and terminography could enrol for postgraduate courses (cf. Alberts 2005c).

The value and need for terminology training is underpinned by the TermTrain initiative of TermNet (an international network for terminology), in cooperation with UNESCO and South African role-players (September 2005 and February 2006). The aim with this terminology training was to emphasise the value of terminology development and training among language practitioners (cf. Alberts 2007).

It is not possible to compile subject-related dictionaries without properly trained terminologists (cf. Alberts 2007).

4.4.2 Value of terminology training and development

4.4.2.1 School of Languages or Language Units at tertiary institutions

Terminology training and the development of term lists could be beneficial to

the school of languages/language unit:

- A school of languages or a language unit would become known as a terminology development centre.
- Draft term lists created by the students/terminologist could be sent to subject specialists, academics, linguists, language practitioners, and language communities for comments, amendments and approval (e.g. NLB approval).
- Verified term lists could be submitted to TCS, NLUs and disseminated to other faculties and departments via the Internet.
- Other faculties and departments would recognise the value of the school of languages/language unit for their own work.
- Schools of languages would enrol more students since a practice-driven course would encourage students (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

4.4.2.2 Other faculties and departments at the specific tertiary institution

Lecturers describe concepts that are usually foreign to students. These concepts are sometimes defined in study material or textbooks but not necessarily in the student's first language. Terminology training and development could, therefore, also be beneficial to other faculties and departments:

- Bilingual and multilingual term lists compiled by the students/terminologist would contain relevant terms and definitions (i.e. excerpted for study guides) and would be disseminated to subject-related departments and faculties.
- Various departments and faculties would become aware of the value of standardised multilingual polythematic terms and definitions.
- Students enrolled in subject-related departments and faculties would be able to understand difficult concepts in their first languages.
- Information retrieval would be easier in the student's first language.
- Even though students have to write their exams in another/second language, they would at least be able to understand the conceptual contents, because the concepts would be clear and internalised.
- Information conveyance and retrieval is paramount in the first language.
- Students would become language conscious and could even supply term lists and glossaries in scripts, dissertations and theses. These term lists could be submitted to the school of languages/language unit at the tertiary institution in a centralised database and disseminated to other stu-

dents. These term lists could also be submitted to the national term bank at TCS, NLS, DAC for (inter)national dissemination (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

4.4.2.3 Value for other faculties and departments at other tertiary institutions

This kind of terminology training and development could also be beneficial to other tertiary institutions that do not present terminology training courses:

- Terminology could be disseminated through virtual networks to other tertiary institutions that could utilise the multilingual terms and definitions.
- Lecturers and students could comment on the available terminology and adjustments could be made.
- Term equivalents and explanations could be supplied in other languages that might not be available yet.
- The terms and definitions could be standardised through frequent usage.
- The terms could be utilised for interpreting and translation services.
- Duplication of subject-related work and the proliferation of terms for the same concept could be avoided (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

4.4.2.4 Value for language offices

Language offices, (non-)governmental language units and private enterprises (e.g. publishing houses) could also benefit from this terminology training and development:

- Terminologists lacking proper training could enrol for terminology courses.
- The basic theories, principles and practices of terminology and terminography are addressed in the terminology course, which decreases the need for in-service training.
- Different types of language practitioners (translators, lexicographers, interpreters, editors, journalists) could enrol for the terminology course seeing that they utilise and need terms in their working environments.
- The different language structures (NLUs, NLBs, TCS, national and provincial language units) could evaluate the course and give feedback to the school of languages/language unit regarding specific needs or content.
- TCS, NLUs and publishing houses would have a list of potential termi-

nologists.

- All these offices could utilise the term lists that are developed.
- The terminology that is created would enhance the potential of the official languages to develop into functional languages in the workplace and into functional languages of science and technology (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

4.4.2.5 Value for Human Language Technologies

Since the term lists would be made available online in electronic format on the Intranet, Internet, etc. there would also be benefits for the national human language technologies virtual network:

- Term lists would be available in electronic format, e.g. at the web pages of tertiary institutions, and could, after the verification and authentication by subject specialists and other experts (e.g. NLBs), become part of the virtual network of the HLT programme.
- As part of the HLT virtual network these multilingual polythematic terms would be available to a variety of users.
- The official South African languages would be in a better position to develop into functional languages of science and technology.
- The official South African languages would also become available and recognised internationally and would even be available on the World Bank of International Terms (WBIT) (cf. Alberts 2007: 8).
- It is noted that Dutch-speaking countries such as the Netherlands and the Flemish-speaking part of Belgium are interested in Afrikaans terminology and Afrikaans terms and explanations would therefore also be internationally available to the Dutch-speaking communities.

4.5 Proposed integrated model

The process described above will train students at tertiary institutions in the basic theory, principles and practice of terminology and terminography and will not only provide trained terminologists to language offices, but would also continuously supply multilingual polythematic terms to various target user communities. This process could enhance the terminology development process.

Activities associated with an integrated terminology training and development model will have to include the following aspects:

- dedicated terminology training, initially undergraduate, but later also postgraduate to all language practitioners-in-training

- practical terminology work that incorporates subject-specific faculties and departments in the development of subject-related terminology lists
- research, especially in terms of the problematic issues relating to the development of terminology in the African languages
- training of lecturers in the theory, principles, practice and management of terminology
- cooperation with the National Language Bodies regarding the language-specific verification and authentication of term equivalents and definitions in the various official languages
- cooperation with the Terminology Coordination Section, National Language Service, Department of Arts and Culture, National Lexicography Units, language units at national and provincial government departments, publishing houses, private initiatives, etc.
- national and international dissemination of multilingual polythematic terminology via the human language technologies virtual network (cf. Alberts 2002; 2004a and b; 2005a, b and c; 2007).

The following integrated model for the development of terminology represents the proposed subject-oriented terminography process by terminology centres at tertiary institutions (cf. Alberts 2005c; Diagram 3):

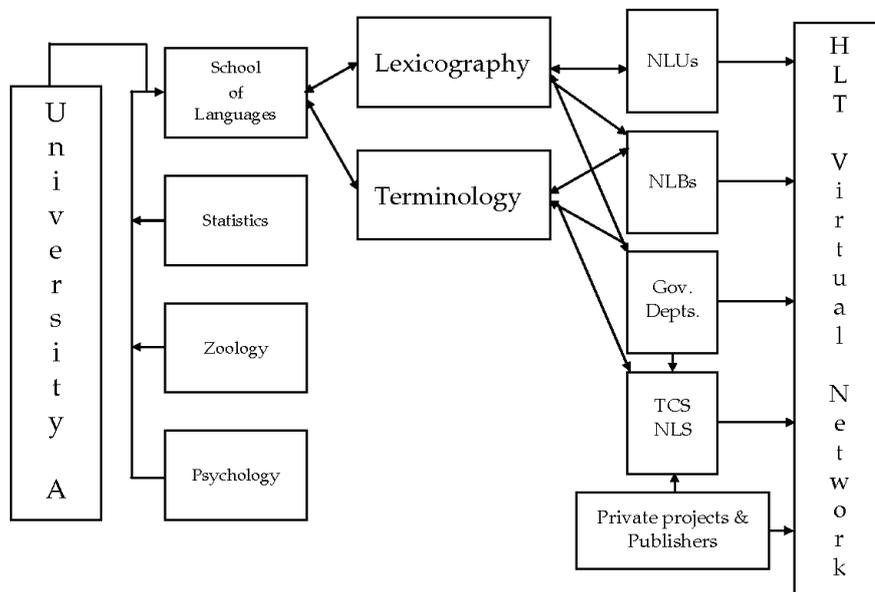


Diagram 3: Subject-oriented terminography process by terminology centres at tertiary institutions

The terminology development process entails the following procedures:

- Term extraction (harvesting/excerption) in the SL from textbooks, tutorials, study guides, dissertations, doctoral theses. These terms should be
 - those with high frequency in the selected texts
 - representing concepts considered critical for students to grasp
 - those which, according to the lecturer, are core terms relating to the subject
 - those which, according to the students, are difficult to grasp in the subject.
- Language students, with the assistance of students majoring in a given subject field, should define the concepts in the SL.
- The definitions should be verified by the language, as well as subject field lecturers.
- The SL terms and definitions are then translated into the TL(s) by language students and verified by the lecturer. Different language departments could assist in the process (i.e. English, Afrikaans, African languages, and even foreign languages such as Roman languages (to determine the etymology of terms)).
- The IT department could assist with database creation, maintenance and data dissemination, i.e. e-learning platforms, Intranet or Internet, where lecturers and students could rate the terms and definitions and make recommendations for improvement.
- The term lists/glossaries should also be presented to the relevant National Language bodies for the verification of language-related principles such as adherence to word-forming principles, spelling and orthographic conventions.
- Term lists or glossaries could be included in tutorials, study guides and textbooks.
- The term lists or glossaries could be disseminated to other tertiary institutions, to the Terminology Coordination Section, other language units, national lexicography units, publishing houses, etc. via the human language virtual network. This would enable penetration of terms in subject fields and in the official languages, which would in the end lead to the standardisation of these terms (cf. Alberts 2005c; 2012a and b).

Although various models were discussed, it should be noted that national collaboration is needed. An integrated model that includes both terminology training and development is therefore proposed. There should be cooperation and interaction and terminological data should be shared. Care should be taken

that there is no duplication of projects. Duplication could result in concepts being denoted in different projects by different SL terms and TL equivalents. This will result in a proliferation of terms denoting the same concepts and it will prevent the standardisation of terminology: this could lead to terminological chaos.

The Department of Higher Education (2010: 8) recognises the cost of developing the official South African languages into languages of instruction and tuition. It also relates the overall student numbers to this developmental issue and proposes regional collaborative frameworks and initiatives (cf. Maseko 2012).

5. Conclusion

It is argued that people understand concepts best in their first language and that it is therefore easier to convey information (i.e. a lecture) through the medium of the first language of both the lecturer and the student. It is essential to internalise information first in the first language, before switching to a second or third language. It is easier to switch to another language (e.g. English) later in life (e.g. working environment) if information was internalised through the first language.

The article focussed on terminology development in tertiary education. Different models for the development of terminology for this educational phase were proposed in order to supply standardised polythematic terminology for tertiary institutions in the various official languages of South Africa. The hegemony of English unfortunately is often the cause why the other official languages struggle with terminology development. The processes described in the three models that culminated in an integrated model will ensure that the official languages will develop into functional languages of higher learning.

The training of terminological theory, principles and practice also received attention. It is argued that schools of languages at tertiary institutions should also present terminology training as an option in language studies. Students following such a course would be suitably trained to join terminology offices and less time would be spent on in-service training. Tertiary institutions would also benefit from terminology courses because students following the course would be able to submit valuable outputs to other faculties and departments. Terminological data could be disseminated to other tertiary institutions and various institutions could collaborate and share data. The virtual human language technologies project could also benefit by receiving multilingual polythematic terminological data on a regular basis and could disseminate the data to subject specialists, linguists, language practitioners and the broader language community.

Both the subject-related language community and the broader language community could benefit from the processes described in the article, since multilingual polythematic terms would be made available for better communi-

cation, minority languages will develop into functional languages in the higher echelons of science, technology, economy, law, etc., and the official South African languages would be globally available as functional (inter)national languages, via the HLT virtual network.

Bibliography

- Alberts, M.** 2002. *Lexicography and Terminology Training at University Level*. Unpublished Document. Pretoria: PanSALB.
- Alberts, M.** 2004a. *Envisaged Lexicography and Terminology Training Centre, Department of African Languages, University of Pretoria, in conjunction with PanSALB*. Unpublished Document. Pretoria: PanSALB.
- Alberts, M.** 2004b. *Establishment of a Lexicography and Terminology Training Centre at the Department of African Languages, School for Languages, University of Pretoria*. Unpublished Document. Pretoria: PanSALB.
- Alberts, M.** 2005a. *Envisaged Lexicography and Terminology Training Centre, Department of African Languages, University of Pretoria, in conjunction with PanSALB*. Unpublished Document. Pretoria: PanSALB.
- Alberts, M.** 2005b. *Training Initiative for Lexicography, Terminology, Human Language Technology and Translation*. Unpublished Document. Pretoria: PanSALB.
- Alberts, M.** 2005c. *Terminologieopleiding op universiteitsvlak*. Paper presented at the SAAWK Symposium, 23–24 June 2005, TE Building, Garsfontein, Pretoria.
- Alberts, M.** 2007. *Voorstel vir die vestiging van 'n Terminologiesentrum aan die Universiteit van Pretoria*. Unpublished document. SAAWK: Pretoria.
- Alberts, M.** 2008. *Centre for Legal Terminology in African Languages (CLTAL)*. Paper read at the PanSALB Conference on Terminology Standardisation, 25–26 September 2008, Gallagher Estate, Midrand.
- Alberts, M.** 2012a. *Terminology as an Aid in Mother-tongue Education*. Paper read at the 17th International Afrilex Conference, 3–5 July 2012, SRC Chambers, University of Pretoria, Pretoria.
- Alberts, M.** 2012b. *Terminology as an Aid in Teaching and Learning*. Paper read at the Closed Workshop on Verification of Scientific Terminology and International Translation Day, 28 September 2012, CPUT, Bellville Main Campus.
- Botha, S., P. le Clus and A.J. Venter.** 2011. *Modern Political Dictionary. Nuwerwetse Politieke Woordeboek*. George: Digi Press.
- Cluver, A.D. de V.** 1982. Aspekte van 'n teorie oor terminologie. *Language Practitioner* 2: 74-90.
- Cluver, A.D. de V.** 1983. *Konsep: Terminologiehandleiding*. Unpublished Document.
- Department of Arts and Culture.** 2011. Presentation to the Portfolio Committee: Arts and Culture on the South African Languages Bill 2011, 16 November 2011. Pretoria: Department of Arts and Culture.
- Department of Higher Education and Training.** 2010. *African Languages Roundtable*. Paper presented at the African Languages Roundtable, 22 October 2010, Unisa, Sunnyside Campus, Pretoria.
- Department of Higher Education and Training.** 2012. *Green Paper for Post-school Education and Training*. Pretoria: Government Printer.

- Department of Sport, Recreation, Arts and Culture.** 2005. *Language Policy Framework of the Gauteng Provincial Government*. Johannesburg: Gauteng Provincial Government.
- Government Gazette.** 2012. Use of Official Languages Act 12 of 2012. *Government Gazette*, Vol. 568, No. 35742, 2 October 2012. Cape Town: The Presidency.
- Government Gazette.** 2013. SA Use of Official Languages Act — Proposed Regulations. *Government Gazette* 26, Notice 411 of 2013, No. 36392, April 2013: 56-67. Cape Town: National Parliament.
- ISO.** 2002. ISO 12616:2002. *Translation-oriented Terminography*. International Organization for Standardization, ISO/TC37/SC2.
- Le Clus, P., S. Botha, A.J. Venter and B. Reyneke.** n.d. *Multilingual Modern Political Dictionary: 1 000 Core Definitions*. Auckland Park: University of Johannesburg.
- Maseko, P.** 2012. *Advancing the Use of African Languages in SA Higher Education: Strategies, Challenges and Opportunities*. Paper read at the Closed Workshop on Verification of Scientific Terminology and International Translation Day, 28 September 2012, CPUT, Bellville Main Campus.
- SAAWK.** n.d. *Bilingual Arts Dictionary. English–Afrikaans/Afrikaans–English*. Unpublished Manuscript.
- Sibula, P.M.** n.d. *Terminology development: Challenges, skills, strategies and stakeholders' role*. Unpublished Paper. Unit for IsiXhosa, Stellenbosh University Language Centre.
- Statistics South Africa.** 2013. *Multilingual Statistical Terminology*. Pretoria: Statistics South Africa.

Internet Reference

<http://www.loc.gov/standards/iso639-2>. Accessed September 2013.

There is No Need for the Terms Polysemy and Homonymy in Lexicography

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

and

Heidi Agerbo, *Centre for Lexicography, Aarhus University,
Denmark (heap@asb.dk)*

Abstract: This paper describes a radically different approach to polysemy and homonymy from the ones normally presented in linguistic and lexicographic literature on this topic. Our main criticism of the traditional approaches lies in their use of the term "word": If a word is defined as a linguistic sign, it means that it only has one expression and one meaning, and this entails that defining polysemy and homonymy as phenomena where one word has two or more meanings is not only problematic — it is impossible. For this reason, we argue that polysemy and homonymy do not exist. Furthermore, we claim that they are not even necessary concepts in lexicography as each lexeme could be represented by a lemma in an information tool. However, by changing the definitions of polysemy and homonymy to phenomena where an expression has two or more meanings, thereby focusing on the expression, it is possible to retain the terms. We propose that the best way to apply and also distinguish between polysemy and homonymy in an information tool would be to present the same expressions with different meanings as well as different grammars as homonyms, while expressions with different meanings but the same grammar are presented as polysems.

Keywords: MEANING, POLYSEMY, HOMONYMY, LEMMA, HEADWORD, LEXICAL WORD, GRAMMATICAL WORD, ORTHOGRAPHICAL WORD, TEXT WORD

Opsomming: Daar is geen behoefte aan die terme "polisemie" en "homonimie" in die leksikografie nie. Hierdie artikel bied 'n radikaal ander benadering tot polisemie en homonimie vergeleke met die benaderings oor hierdie onderwerp wat tipies in taalkundige en leksikografiese literatuur te vinde is. Ons belangrikste kritiek op die tradisionele benaderings lê in die gebruik van die term "woord". As 'n woord gedefinieer word as 'n linguïstiese teken beteken dit dat dit net een uitdrukkingsvorm en een betekenis het. Dit impliseer dat dit nie problematies is nie maar onmoontlik om polisemie en homonimie te omskryf as verskynsels waar een woord twee of meer betekenis het. Om hierdie rede voer ons aan dat polisemie en homonimie nie bestaan nie. Daarbenewens stel ons dit dat hulle nie eers nodige begrippe in die leksikografie is nie aangesien elke lekseem deur 'n lemma in 'n inligtingswerktuig verteenwoordig kan word. Deur die definisies van polisemie en homonimie te verander tot verskynsels waar 'n uitdrukking twee of meer betekenis het, is die fokus op die uitdrukking en is dit moontlik om die terme te behou. Ons stel voor

dat die beste manier om tussen polisemie en homonimie te onderskei en om dit toe te pas in 'n inligtingswerktuig is om dieselfde uitdrukkings wat verskillende betekenis en verskillende grammatiese waardes het as homonieme aan te bied terwyl uitdrukkings met verskillende betekenis maar dieselfde grammatiese waarde as poliseme aan te bied.

Sleutelwoorde: BETEKENIS, GRAMMATIESE WOORD, HOMONIMIE, LEKSIKALE WOORD, LEMMA, ORTOGRAFIESE WOORD, POLISEMIE, TEKSWOORD, TREFWOORD

1. The tradition in linguistics and lexicography

In traditional linguistic and lexicographic literature, polysemy and homonymy are gathered under the umbrella term ambiguity. The common and shared definition of polysemy is that it is the phenomenon where a single word has two or more related meanings, and the common and shared definition of homonymy is the phenomenon where a word has two or more unrelated meanings. Another typical definition of homonymy is that it is the phenomenon where two or more words that are identical in form have different meanings. There are two main problems with these definitions.

The first one is the foundation of the distinction made between polysemy and homonymy. The typical one, as already mentioned, is the semantic criterion. Based on this criterion, word meanings are analysed as either semantically related or unrelated, which is highly based on individual intuition. Some typical examples of this kind of principle are discussed in Geeraerts (2013: 492-494) for *primer* and *fresh*, though he does not use the term "homonym", only the terms "polysem", "sense" and "subsense":

Lemma 1 with expression 1 + meaning A

Lemma 2 with expression 2 + the common meaning element a as homonym I

polysem 1 for meaning B
polysem 2 for meaning C
polysem 3 for meaning D

Lemma 3 with expression 2 + the common meaning element b as homonym II

polysem 1 for meaning E
polysem 2 for meaning F
polysem 3 for meaning G

Model (1): A polysem-homonym model based on semantic distinctions

If a word is homonymous, the homonyms are presented as individual dictionary entries, but if it is polysemous, the polysems are grouped under the same entry word. Establishing whether two meanings of a word are homonymous or polysemous (or even subpolysemous), is, however, a task that will normally

result in different solutions from one lexicographer to the next.

A different type of distinction between homonymy and polysemy is presented from the etymological approach, which looks at meaning from a diachronic perspective, in which two meanings are unrelated if it is impossible to trace any historical connection between them. This approach, however, is useless from a synchronic perspective, because even though two senses of a word are synchronically unrelated, they may be diachronically related.

A third approach is the formal approach with which polysemy and homonymy are established based on formal criteria. This means that words that are orthographically similar but have different inflectional paradigms (also within the same part of speech) are defined as homonyms, whereas orthographically similar words belonging to the same part of speech and with the same inflectional paradigm are defined as polysems.

These three as well as other approaches applied in the work on polysemy and homonymy overlook one important thing that leads to the second critical point and major problem in the literature on polysemy and homonymy: The use of the term "word".

2. The linguistic sign and types of words

In the structuralist tradition of Saussure and Hjelmslev, one of the main notions is the linguistic sign, which is defined as containing two parts: the expression and the content, with a mutual relationship between the two. Hjelmslev calls this relationship solidarity:

Saaledes bestaar der ogsaa solidaritet mellem tegnfunktionen og dens to funktiver: udtryk og indhold. Der vil aldrig foreligge en tegnfunktion uden at begge disse funktiver samtidig er til stede. (Hjelmslev 1943: 45)

[Eng.: Thus, solidarity also exists between the sign and its two functives: expression and content. There will never be a sign without the mutual presence of both functives.]

When linguists and lexicographers talk about polysemy, they typically say something along the line of polysemy — and also homonymy — is when one word has more than one meaning. This is a very superficial definition, but that is not the main problem. The problem is the use of "word". Based on the quote by Hjelmslev, a "word" cannot be a linguistic sign because according to the definition of the sign, we have two different linguistic signs when there are two different meanings:

for expression 1 with solidarity to content A, we have one sign

for expression 1 with solidarity to content B, we have another sign

The problem that must be solved in order to understand what is meant by polysemy and homonymy is to find out what we mean when we use the term "word".

A word is not a word, or, more specifically, we have different kinds of words (Bergenholtz and Mugdan 1979: 12-29 and 116-125) which are: orthographical word, text word, grammatical word, lexical word (lexeme) and dictionary word (lemma). These word types are all relevant in lexicography:

An **orthographical word** is defined as a sequence of letters between blanks and sentence signs (like comma). Orthographical words are also used as search strings in e-dictionaries. But an orthographical word is not a linguistic sign because one and the same orthographical word can be related to different contents.

A **text word** is a concrete word in a text with a specific spelling, meaning, grammar etc. If - especially in the case of a reception problem - a potential dictionary user comes across a text word in the text he is reading, he will write an orthographical word corresponding to the expression of the text word in the search field or look for it in the lemma list. A text word is a linguistic sign and as such the main object for dictionaries, but it is not used, with the exception of author's dictionaries, directly in the production and the use of dictionaries because text words are unique in the specific texts in which they occur, and the search in dictionaries is conducted using an orthographic word.

A **grammatical word** is an expression with at least one nucleus morpheme and for adverbs, verbs and nouns also at least one grammatical morpheme (Bergenholtz and Mugdan 1979: 118-125). A grammatical word belongs to a certain inflexion paradigm (for words without inflexion, this paradigm consists of only one grammatical word). All or some of the grammatical words belonging to one lexeme are listed in many dictionaries after the lemma. It is possible to search for them in many e-dictionaries if searching with a search string equal to this grammatical word, i.e. not only the typical lemma form. A grammatical word is not a linguistic sign because one and the same grammatical word can belong to more than one lexeme.

A **lexical word** or a **lexeme** is the abstraction for an amount of grammatical words belonging to the same stem and the same inflexion paradigm. In order to show the abstract status of a lexical word, Bergenholtz and Mugdan (1979) gives lexical and grammatical words different notions: BUCH_{Substantiv} {Buch, Buch, Buches, Buch, Bücher, Bücher, Bücher, Büchern}. A lexical word is a linguistic sign, and different stems result in different lexical words, hence we also define a stem as a linguistic sign with an expression and a meaning. Stems include all alloforms, as for Buch- and Büch-, but the main point is that we do not have alloforms for the content. If we have different stem meanings, then we have different linguistic signs. For instance, in Bergenholtz and Agerbo (2014) we show that the stem *pigtail* is connected to at least six different lexical words with six very different meanings.

A **dictionary word**, **headword** or **lemma** is the abstraction for an amount of grammatical words, but it is not the same as a lexical word, because, contrary to lexical words, different stem meanings do not result in different headwords. In paper dictionaries, a lemma is normally shown as the expression of

one and only one of the grammatical words belonging to an inflexion paradigm with the same stem. In dictionary articles in e-dictionaries, the user will often only see the expression for one of the grammatical words in the inflexion paradigm, but the search algorithm can search in the whole paradigm.

In the lexicographical tradition, described in this article as polysem-homonym-models (1) and (3), a lemma is not a linguistic sign because a lemma can represent different lexical words (sometimes it represents only one lexeme, in other cases it represents several lexemes). Hence, there is no solidarity between one expression and one content. Only in the radical solution, which has been described below as model (2), where we discard polysemy and homonymy and instead connect each lexical word to its own lemma, the lemma could be defined as a linguistic sign.

3. Polysemy and homonymy do not exist

According to the argument presented in Chapter 2, we cannot speak about polysemy and homonymy if we relate these terms to the linguistic sign. However, in lexicography we can do well without these terms. The solution to get around them would be to have one lemma (one dictionary word) for each lexeme and provide the grammar items for each lemma. This means that we have one and only one lexeme in each dictionary article. This could easily be done in e-dictionaries, though in printed dictionaries this solution would require too much space for at least one-volume dictionaries.

Lemma 1 with expression 1 + grammar Z + meaning A

Lemma 2 with expression 2 + grammar Z + meaning B

Lemma 3 with expression 2 + grammar Z + meaning C

Lemma 4 with expression 2 + grammar Z + meaning D

Lemma 5 with expression 2 + grammar Z + meaning E

Lemma 6 with expression 2 + grammar Y + meaning F

Lemma 7 with expression 2 + grammar Y + meaning G

Model (2): A model without polysemy and homonymy

In the case of lemma 1 with expression 1, which is related to only meaning A and grammar Z, of course, we cannot talk about polysemy or homonymy because there is only one meaning. But neither can we call the lemma a linguistic sign. Imagine that we have expression 2 that is related to more than one meaning; for each lemma, we find partly the same, partly not the same grammar, but in each dictionary article, there are different meanings. In this case, the lemma is not a linguistic sign. Though expression 1 is connected to only one meaning and one grammar, it does not make sense to call the lemma a linguistic sign because how, then, would we distinguish between lemmas that are linguistic signs and lemmas that are not?

4. Polysemy and homonymy do exist

In Chapter 3 and in Bergenholtz and Agerbo (2014), we argue that polysemy does not exist and that we do not consider polysemy a useful term in lexicography. However, we could also move in the opposite direction and argue that both polysemy and homonymy actually do exist. However, this requires a different definition of the terms. If we focus on the expression of the linguistic sign, i.e. if polysemy is related to the expression of the linguistic sign instead of the linguistic sign with one expression and one meaning, it is possible to argue in favour of both polysemy and homonymy. With this definition, it is possible to say that polysemy is the phenomenon where the same expression (not word!) is connected to several meanings. Hence, an expression is polysemous if it is connected to more than one meaning. If there are two or more meanings and a distinction appears in the part of speech or the inflection paradigm within the same part of speech, the expression is not polysemous, but homonymous.

This approach is contrary to the typical one in linguistics and lexicography, though dictionaries such as The Danish Internet Dictionary (2014) and Wahrig (1986) already apply this approach to polysemy and homonymy. Of course, this entails that some of the homonymous meanings are semantically closer connected than others, but this approach instead has the advantage that polysemy and homonymy may be treated systematically instead of intuitively. Whether the homonymous meanings or the polysemous meanings connected to an expression are close or not may be a relevant linguistic problem to explore, but for the dictionary user this is unimportant, at least in a reception situation. It could be relevant in a cognitive dictionary, if the user is interested in learning more about the expression.

Using the same distribution of expressions and meanings as in model (2), we now only have 3 lemmas instead of 7:

Lemma 1 with expression 1 + grammar Z + meaning A

Lemma 2 with expression 2 + grammar Z as homonym I

polysem 1 for meaning B

polysem 2 for meaning C

polysem 3 for meaning D

polysem 4 for meaning E

Lemma 3 with expression 2 + grammar Y as homonym II

polysem 1 for meaning F

polysem 2 for meaning G

Model (3): A polysem-homonym model based on grammatical distinctions

5. Application of the three models

With model (1), we can use a diachronic argumentation intended for a cogni-

tive dictionary to show that a certain meaning is the original one (the earliest one) and that it forms the basis for later derived meanings of a certain orthographic word. As an example, we can take the lemma *pigtail*, for which the meaning 'the hindmost part of a pig' corresponds to the earliest use of the word (see Bergenholtz and Agerbo 2014). According to model (1), the six meanings expressed by the lemma *pigtail* are not homonyms, but polysems:

1. the small, curly body part on the hindmost part of a pig
2. the fleshy tail from a pig used for cooking, e.g. in soups, or prepared as a snack for dogs
3. a length of hair that is tied at the back of the head or at each side of the head, sometimes in a braid; mainly worn by women
4. a waist-long braided ponytail made by braiding the hair at the back of the neck, often also combined with the hair on the front of the head being shaved off, which men in ancient China were demanded to wear as a symbol of submission and obedience
5. item of artificial material resembling braided human hair, which is attached to the back of a Chinese hat and often worn as part of a costume imitating a Chinese man from ancient China
6. one, two or several pom-poms or strings of yarn or other fabric attached to a knitted hat, which often resemble a braid of hair or the tail of a pig

Instead of this presentation, model (1) could be based on a synchronic description of the use of this orthographic word. Applying model (1), this would result in three homonyms each with two polysems. This presentation would also be suitable for a dictionary with cognitive functions:

Homonym I 'the hindmost part of a pig'

1. the small, curly body part on the hindmost part of a pig
2. the fleshy tail from a pig used for cooking, e.g. in soups, or prepared as a snack for dogs

Homonym II 'natural or artificial hair'

1. a length of hair that is tied at the back of the head or at each side of the head, sometimes in a braid; mainly worn by women
2. a waist-long braided ponytail made by braiding the hair at the back of the neck, often also combined with the hair on the front of the head being shaved off, which men in ancient China were demanded to wear as a symbol of submission and obedience

Homonym III 'hair-like decoration attached to a hat'

1. item of artificial material resembling braided human hair, which is attached to the back of a Chinese hat and often worn as part of a costume imitating a Chinese man from ancient China
2. one, two or several pom-poms or strings of yarn or other fabric attached to a knitted hat, which often resemble a braid of hair or the tail of a pig

For communicative functions, either model (2) or (3) would be the most suitable one as it will always be the specific use of a word that forms the point of departure in case of a reception problem or a text production problem. First and foremost, the user needs to know what the word means in a specific context (reception) or whether it is possible to use a certain word with the meaning the user wishes to express (production). The only difference between the outcomes of models (2) and (3) for this example is that model (3) results in one lemma with one grammatical description, whereas model (2) results in six different lemmas with each their grammatical description. The solutions based on models (1) and (3) are identical for *pigtail* as for this lemma there are no grammatical distinctions between the different uses of the word. Had a difference in the grammar been the case, model (1) would still be without homonymy, but model (3) would include homonymy.

6. Which model can be recommended for modern lexicography?

The suggestions provided above apply both to monolingual and bilingual dictionaries; there are no significant differences. This is also the case when a bilingual dictionary contains no definitions because equivalents are always selected based on the definitions that would have been selected in a bilingual dictionary with definitions. In many of the dictionaries that we have helped produce, e.g. The Danish Internet Dictionary (2014) or The Spanish Accounting Dictionary (2014), we have applied model (3). This is a clear and precise model, see also Tarp (2008). With this model, contrary to model (1), all lexicographers will make the same labelling of senses as either homonymous or polysemous. Similarly, the principles for model (3) can easily be explained to the dictionary users. We do acknowledge, however, that even though it will be fairly easy for users to understand that different parts of speech result in different grammars, it may be difficult for them to understand the difference between two inflectional paradigms within the same part of speech. The clearest and simplest model is therefore model (2). This one is easier for lexicographers to apply in the production of dictionary articles, and it would also be easier to explain to the users how they can find the pieces of information they are looking for. Neither model (2) nor (3) is connected to any theoretical contradictions, which is the major problem for model (1). What speaks in favour of model (3) is first and foremost that this one is closer to the solution that dictionary users are familiar with.

7. References

- Bergenholtz, H. and H. Agerbo. 2014. Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries. *Hermes. Journal of Language and Communication in Business* 52: 125-139.
- Bergenholtz, H. and J. Mugdan. 1979. *Einführung in die Morphologie*. Stuttgart: Kohlhammer.

- Geeraerts, D.** 2013. The Treatment of Meaning in Dictionaries and Prototype Theory. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography*: 487-495 Berlin: Mouton de Gruyter.
- Hjelmlev, L.** 1943. Omkring sprogteoriens grundlæggelse. *Festskrift udgivet af Københavns Universitet i anledning af Universitetets Aarsfest, November 1943*: 1-113. Copenhagen: Bianco Lunos Bogtrykkeri.
- Tarp, S.** 2008. Lexicography and the Linguistic Concepts of Homonymy and Polysemy. *Lexicographica* 17: 22-39.
- The Danish Internet Dictionary.** 2014. = H. Bergenholtz in collaboration with Agerbo, H. in cooperation with Bodilsen, A., S.W. Dohrn, J. Esmann, H.R. Gudmann, Chr. Kjølhed, J. Nguyen, M.L.K. Nielsen, H. Thers. Database: Almind, R. and M. Carlsen. *Den Danske Netordbog*. Odense: Ordbogen.com.
- The Spanish Accounting Dictionary.** 2014. = Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P.G. Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á.S. Ruano, S. Tarp and M. Velasco Sacristán. 2014. *Diccionario Español de Contabilidad: Recepción*. Base de Datos y Diseño: R. Almind and J.S. Nielsen. Hamburg: Lemma.com.
- Wahrig = Wahrig, G.** 1986. *Deutsches Wörterbuch. Mit einem „Lexikon der deutschen Sprachlehre“*. München: Mosaik Verlag.

Explaining Dysfunctional Effects of Lexicographical Communication*

Herman L. Beyer, *Department of Language and Literature Studies,
University of Namibia, Windhoek, Namibia (hbeyer@unam.na)*

Abstract: During a keynote address at an international conference of AFRILEX articles from a bilingual dictionary with Afrikaans as one of the treated languages was presented. One of the articles was that of the lemma **vuvuzela**, which contained the example sentence *Vuvuzelas maak 'n groot lawaai by sokkerwedstryde* [± Vuvuzelas make a lot of noise at soccer matches]. A member of the audience criticised this example for apparently not reflecting the notion of festivity and celebration, with which the vuvuzela is also associated; instead, it seemed that the example focused only on a negative feature of the vuvuzela. From the ensuing discussion it became clear that there seemed to be no theoretical framework against which the criticism could be validated and productively dealt with, even though the lexicographer ultimately offered to review the example.

This article introduces elements of the *theory of lexicographical communication* and applies them to scaffold such a framework. It is argued that indicators in dictionary articles can be regarded as lexicographic utterances that carry various types of lexicographic messages. These can be systematically and formally analysed to identify functional, non-functional and dysfunctional effects of lexicographical communication. Problems with lexicographical communication can then be diagnosed and addressed. This potential is illustrated by treating the above-mentioned occurrence as a case study. In conclusion, the value of the relevant elements of the theory for the evaluation of dictionaries is briefly outlined.

Keywords: APPEAL, DICTIONARY, DYSFUNCTIONAL EFFECT, EXPRESSIVE, FUNCTION, FUNCTIONAL EFFECT, INFORMATION, LEXICOGRAPHER, LEXICOGRAPHIC MESSAGE, LEXICOGRAPHIC UTTERANCE, LEXICOGRAPHICAL COMMUNICATION, LEXICOGRAPHY, NON-FUNCTIONAL EFFECT, REFERENTIAL, RELATIONAL, TARGET USER

Opsomming: Die verklaring van disfunksionele effekte van leksikografiese kommunikasie. Gedurende 'n hoofreferaat by 'n internasionale konferensie van AFRILEX is artikels van 'n tweetalige woordeboek met Afrikaans as een van die behandelde tale aangebied. Een van die artikels was dié van die lemma **vuvuzela**, wat die voorbeeldsin *Vuvuzelas maak 'n groot lawaai by sokkerwedstryde* bevat. 'n Lid van die gehoor het hierdie voorbeeld gekritiseer omdat volgens hom dit nie die idee van feestelike weerspieël waarmee die vuvuzela ook geassosieer word nie; dit het naamlik gelyk of die voorbeeldsin op slegs 'n negatiewe aspek van die vuvuzela fokus. Uit die voortspruitende bespreking het dit geblyk dat daar nie 'n teoretiese raamwerk

* An earlier version of this article was presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1-3 July 2014.

bestaan waarvolgens die kritiek bekragtig en produktief hanteer kon word nie, afgesien daarvan dat die leksikograaf aangebied het om die voorbeeldsin te hersien.

Hierdie artikel stel elemente van die *teorie van leksikografiese kommunikasie* bekend en pas hulle toe om so 'n raamwerk te skep. Daar word aangevoer dat aanduiders in woordeboekartikels beskou kan word as leksikografiese uitings wat verskillende tipes leksikografiese boodskappe dra. Hulle kan sistematies en formeel analiseer word om funksionele, niefunksionele en disfunksionele effekte van leksikografiese kommunikasie te identifiseer. Probleme met leksikografiese kommunikasie kan dan gediagnoseer en opgelos word. Hierdie potensiaal word geïllustreer deur die toepassing van die teorie op die bogenoemde voorval as 'n gevallestudie. Ter afsluiting word die relevante elemente van die teorie se waarde in die evaluering van woordeboeke kortliks uiteengesit.

Slutelwoorde: APPÈL, DISFUNKSIONELE EFFEK, EKSPRESSIEF, FUNKSIE, FUNKSIONELE EFFEK, INLIGTING, LEKSIKOGRAAF, LEKSIKOGRAFIE, LEKSIKOGRAFIESE BOODSKAP, LEKSIKOGRAFIESE KOMMUNIKASIE, LEKSIKOGRAFIESE UTTING, NIEFUNKSIONELE EFFEK, REFERENSIEEL, RELASIONEEL, TEIKENGEBRUIKER, WOORDEBOEK

1. Introduction

This article was prompted by an occurrence during question time after one of the keynote addresses at the annual international conference of the African Association for Lexicography (AFRILEX) held in Port Elizabeth, South Africa, from 2 to 5 July 2013. The keynote address in question dealt with a bilingual dictionary that is being compiled with Afrikaans as one of the treated languages. One of the aspects touched upon during the address was example phrases and sentences. In this regard the article of the lemma **vuvuzela**, *inter alia*, was presented. This article contained the following Afrikaans example sentence: *Vuvuzelas maak 'n groot lawaai by sokkerwedstryde* [\pm Vuvuzelas make a lot of noise at soccer matches]. In a comment during question and discussion time one delegate expressed disappointment at the fact that he could not observe the notion of festivity and celebration, with which the vuvuzela is also associated, in the example. It seemed regrettable to him that the vuvuzela is associated only with excessive noise in the dictionary article. In response to this comment, the keynote speaker and some members of the audience argued that the example sentence was included due to the relatively high co-occurrence of the type *vuvuzela* with the type *noise* in the relevant corpus. However, this line of defence did not seem to satisfy the objection. Ultimately, the keynote speaker offered that the dictionary article in question could eventually be revisited and revised to address the criticism.

What became clear during the discussion, however, was that there seemed to be no metalexicographical framework against which this type of criticism could be validated and dealt with productively. The aim of this article is to address this research problem by tentatively offering such a framework. This proposed solution falls within the parameters of a *theory of lexicographical communication*.

Firstly, the elements of the theory of lexicographical communication that scaffold the proposed framework will be developed, after which the focus will fall on the research problem itself. In conclusion the value of the framework for the formative and summative evaluation of dictionaries will be outlined in the form of a set of general guiding questions.

The theory that is introduced in this article allows for formal representations of propositions and qualities, as will become apparent. However, most of the formal representations are accompanied by natural language equivalents and therefore the article can be read without having to interpret each formal representation.

It should be stated at this point that the purpose of this article and the theory presented in it is not to choose any side in the apparent disagreement between the lexicographer and the delegate, nor is it to resolve any issue surrounding the vuvuzela. The research problem is the absence of a scientific framework for discussion. The purpose of the study is to provide a scientific framework which, it is hoped, could facilitate fruitful discussion in addressing issues like the one raised above.

2. The theory of lexicographical communication

The theory of lexicographical communication (= TLC) can be regarded as an ongoing study to determine to what extent a specific communicative perspective might facilitate new and ideally innovative theorising in lexicography. A systematic exposition of the foundations of this approach has not been published as yet, but is being prepared: Beyer (in preparation). However, some preliminary elements have been tentatively and informally applied in Beyer (2009, 2010, 2011 and 2013).

With a "communicative perspective" is meant an approach originating from communication theory. A basic definition of the core concept *communication* is therefore in order for the purposes of this article (translated and slightly adapted from Van Cuilenburg et al. 1992: 10):

- (1) *Communication* is a process during which a sender conveys information via a medium to a receiver with the goal of achieving a specific effect with the receiver.

The concept *lexicographical communication* refers to a specialised type of communication limited to specific types of information (i.e. about lexical items) and a specific type of medium (i.e. the dictionary) for specific types of effects. In lexicographical communication the lexicographer usually plays the role of the sender (in so far as the concept *lexicographer* can be defined – cf. Gouws 2012) and the target user plays the role of the receiver. The concept *information* will be dealt with in section 3 below.

The TLC is being developed from the lexicographical communication

model first proposed by Beyer (2006) in an effort to present a framework for the systematic description of user-friendliness in dictionaries. Subsequently, Yong and Peng (2007) independently proposed a communicative approach to bilingual lexicography.¹ Ptaszyński (2009: 212), in reviewing the work, refers to it interchangeably as "the communicative theory of lexicography" and "the communication theory of lexicography". Although Yong and Peng's work is based on an alternative communication model and consequently differs almost fundamentally from the current TLC, it served as an inspiration to further explore the potential of a communication model for theory development in lexicography.²

Since Beyer (2006) the model has evolved somewhat but has not changed fundamentally. It can now be presented in figure 1:

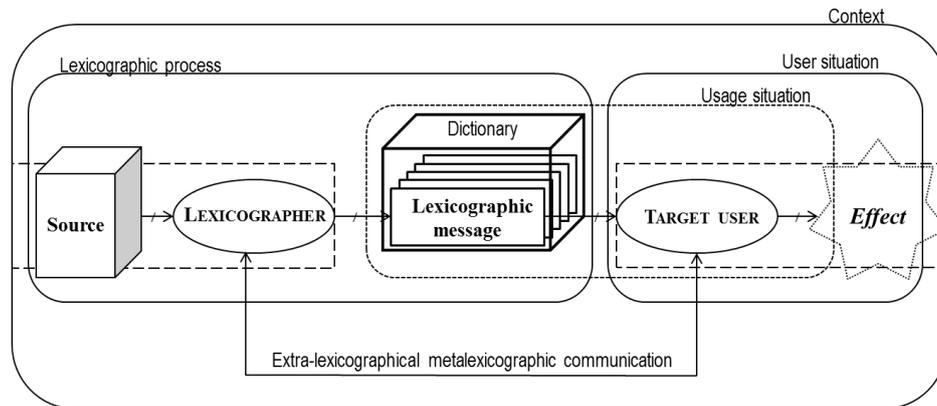


Figure 1: A lexicographical communication model

In this model a number of components of the process of *lexicographical communication* can be distinguished and described, as is done preliminarily in Beyer (2006). Two of these components are the primary components or immediate constituents of this process, namely the lexicographic process on the one hand and the dictionary consultation process (i.e. the user situation in fig. 1) on the other. They overlap to both enclose the lexicographic message and its container, the dictionary as the medium of lexicographical communication. The other components, e.g. the source, the lexicographer, the target user and the user situation, are related to either of the primary components. Even though the components can be individually identified and described, the model illustrates their mutual dependency and necessity: lexicographical communication cannot take place if one of them is missing. The only component that overarches the primary components is the context component, which represents variables that are not part of the communication process per se, but determine

the values of at least a subset of the variables in all other components while the reverse is not applicable to the same extent. This component will be elaborated on in section 6.3.

In reading the lexicographical communication model above, the TLC can be summarised as describing a sub-process, the lexicographic process, during which the lexicographer sources, selects and encodes lexicographical information in a lexicographic text segment, followed by a second sub-process, the dictionary consultation process, during which the dictionary user decodes and interprets the lexicographic message in the usage situation. This is followed by the effect of the communication process, when the user applies the meaning of the lexicographic message to the user situation. Both sub-processes and their effect occur in a particular context (cf. 6.3). At the centre are the lexicographic message and its container, the dictionary.

Cruse (2011: 5) presents the following stages in a typical act of *linguistic communication*, each of which he subsequently discusses in more detail:

- (2) a. The speaker normally has a purpose in communicating.
- b. The speaker constructs a message to be communicated.
- c. The speaker constructs an utterance with which to convey the message.
- d. The speaker transforms the utterance into a physical signal.
- e. The speaker transmits the signal.
- f. The addressee receives the signal.
- g. The addressee decodes the signal to recover the utterance.
- h. The addressee reconstructs the message from the utterance.
- i. The addressee infers the purpose of the communication.

The strong identity between the process of lexicographical communication and linguistic communication is obvious and not surprising, since language is a code system used for communication. Contemporary lexicographical theories emphasise the view that lexicography is not a branch of linguistics or applied linguistics (cf. e.g. Tarp 2008), although this view is not universally accepted (cf. e.g. Atkins and Rundell 2008, Fontenelle 2011). Whichever opinion applies, the value of linguistics in practical and theoretical lexicography should be acknowledged, simply because the code system generally employed for lexicographical communication is language, as shown above. It could be argued that linguistic theory might perhaps not always be best suited for explaining *why* and *what types* of information should be encoded in a given type of dictionary and *what structures* should frame the lexicographic utterances, but it surely has the potential to contribute to explaining *how* lexicographic messages are encoded in lexicographic utterances, *how* the utterances are decoded and interpreted, and *how* the target user applies the derived information. (Cf. also Bogaards 2010.)

In a review of Tarp (2008) which concludes that the function theory fails as a theory, Piotrowski (2009: 485) states in closing:

[A] new theory of lexicography would be one that would take seriously what we know now about pragmatics and discourse, text structure, [sic] and would account for the contribution of particular textual elements to the dynamic meaning of a text ...

Taking Piotrowski's cue and considering the views of other critics of the function theory (cf. Bogaards 2010, De Schryver 2012, Rundell 2012), the potential of the discipline of pragmatics in the development of the TLC is currently being investigated. Of particular interest seems to be *speech act theory*, which was first proposed by Austin (1962) and further developed by Searle (1969, etc.), and which is in fact considered as a branch of the theory of communication (Kempson 1977, Bierwisch 1980). The basic elements of speech act theory are applied later in this article (cf. 4.2.4). Another theory in pragmatics that seems promising for application to lexicographical communication and accommodation in the TLC is the theory of *conversational implicature* (Grice 1975, etc.). Levinson (1983: 101-102) summarises Grice's co-operative principle and maxims with regard to the efficient and effective use of language (i.e. communication), which would clearly also resonate strongly with the lexicographer, as follows:

- (3) a. *The co-operative principle*
Make your contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.
- b. *The maxim of Quality*
Try to make your contribution one that is true, specifically:
 - (i) do not say what you believe to be false;
 - (ii) do not say that for which you lack adequate evidence.
- c. *The maxim of Quantity*
 - (i) Make your contribution as informative as is required for the current purposes of the exchange.
 - (ii) Do not make your contribution more informative than is required.
- d. *The maxim of Relevance*
Make your contributions relevant.
- e. *The maxim of Manner*
Be perspicuous, and specifically:
 - (i) avoid obscurity;
 - (ii) avoid ambiguity;
 - (iii) be brief;
 - (iv) be orderly.

Furthermore, the TLC emphasises the textual nature of lexicographical communication. As such, it explores, in addition to existing lexicographical and communication theory, disciplines like discourse analysis, text linguistics and document design in attempting to describe lexicographical communication.

The focus now turns to lexicographic messages in approaching the research problem.

3. Lexicographic messages

As stated above, communication is a process of conveying information. Following Cleary (2003: 5) and Steinberg (2007: 45-46), the concept of *information* in human communication can be described as follows:

- (4) *Information* is the factual knowledge, ideas, thoughts and feelings that can be conveyed by means of communication.

In order for the information to be conveyed, it first has to be constructed into a message, which can then be encoded into an utterance (following Cruse's stages of linguistic communication). From this a definition for the concept *message* can be derived logically:

- (5) A *message* is a unit of information that can be encoded in an utterance.

An *utterance*, in turn, is a discrete unit of communication that consists of a sign or an ordered string of signs forming symbols (in terms of linguistic communication: sounds and letters forming lexical units) used by the sender to encode a message according to the code (the grammar) of the code system (the language) that is used to communicate.

In the literature on general communication theory a clear distinction between *message* and *utterance* is not often formulated, unlike for linguistic communication by Cruse in (2) above. In some cases these terms would ostensibly be regarded as referential synonyms; compare for example Sebeok (2006: 44), who states that "[a] message is a *sign*, or consists of a string of signs". For this reason the definition of the concept *message* in (5) has been crafted for the purposes of the TLC and not quoted from existing communication theory literature. A further motivation for this step lies in the *principle of expressibility*, introduced by Searle (1969: 19) in speech act theory, which states that "whatever can be meant can be said." This principle is central to the TLC as it currently stands.³

A clear distinction between *message* and *utterance* is not only valid, but necessary, as more than one message can be encoded in an utterance. In fact, it will be shown in section 4 that at least four types of messages are encoded in an utterance.

In lexicographical communication, then, specific types of information (i.e. about lexical items) are constructed in specific types of messages that are encoded in specific types of utterances (i.e. lexicographic text segments) during the lexicographic process. During the dictionary consultation process, the target user locates and decodes the utterances to reconstruct the messages and derive the information, which is interpreted against the user's frame of refer-

ence and then applied to the user situation.

To illustrate these terms, the following partial dictionary article will serve as example:

(6) **book**, *n.*

This partial article consists of three lexicographic utterances LU_1 , LU_2 and LU_3 , in the form of separate (uncategorised) text segments:

- (7) a. $LU_1 = \text{"book"}$
b. $LU_2 \setminus LU_1 = \text{","}$
c. $LU_3 \setminus LU_1 = \text{"n."}$

("LU_x\LU_y" reads "LU_x, as addressed at LU_y."

Lexicographer *l* wants to convey information i_l , that the lexical item *book* is a noun, to target user *u*. The lexicographer constructs this information in lexicographic message LM_l :

(8) $LM_l = \text{The lexical item } book \text{ is a noun.}$

Therefore, i_l is constructed in LM_l for the lexicographer:

(9) $i_l[LM_l]$

("x[y]" reads "x is constructed/encoded in y or reconstructed/decoded from y".)

This message is then encoded in lexicographic utterance $LU_3 \setminus LU_1$ in the partial dictionary article above:

(10) $LU_3 \setminus LU_1 = \text{"n."}$

Therefore, LM_l is encoded in $LU_3 \setminus LU_1$ for the lexicographer:⁴

(11) $LM_l[LU_3 \setminus LU_1]$

Therefore:

(12) $i_l[LM_l[LU_3 \setminus LU_1]]$

Target user *u* locates $LU_3 \setminus LU_1$, decodes it and reconstructs LM_u :

(13) $LM_u = \text{The lexical item } book \text{ is a noun.}$

Therefore, LM_u is decoded from $LU_3 \setminus LU_1$ for target user *u*:

(14) $LM_u[LU_3 \setminus LU_1]$

Target user *u* reconstructs from LM_u information i_u , that the lexical item *book* is a noun, and then applies this information to user situation s_u , which constitutes

effect E_u . Therefore:

$$(15) \quad i_u[LM_u[LU_3 \setminus LU_1]] \rightarrow E_u$$

("x → y" reads "From x follows (naturally) consequence y".)

The sequence of lexicographical communication is therefore:

$$(16) \quad i_l \rightarrow LM_l \rightarrow LU_3 \setminus LU_1 \rightarrow LM_u \rightarrow i_u \rightarrow E_u$$

This simple example represents the ideal case in user-centred lexicography: $i_l = i_u$, which satisfies user need n_u in user situation $s_{u'}$ as is evidenced by effect E_u .

Generally speaking, contemporary lexicographical theories use the term *data* to refer to text segments and *information* to refer to the "information" that the target user derives from decoding the text segments (cf. Tarp 2008: 97-100). It would seem, however, that *data* and *information* in these theories refer almost exclusively to the relevant facts about the treated lexical items. This view seems to coincide with the somewhat limited view of the functions of language that Brown and Yule (1983: 2) claim linguists and linguistic philosophers generally adopt, namely "that the most important function is the communication of [data]". However, as the definition of the term *information* in communicative terms above suggests, the concept encapsulates more than factual knowledge and can crucially also include informational dimensions like attitudes, emotions, etc. that cannot necessarily be expressed in terms of data. It seems, then, that the denotation of the term *information* is slightly broader in the TLC than in contemporary lexicographical theories.⁵

Nevertheless, the TLC may not ignore the valid rationale for the distinction between *data* and *information* in these theories, one that is also supported by communication theory (albeit without using the same terms). For this reason, information constructed in an LM_l by the lexicographer is referred to above as "information_l" (= i_l), and information reconstructed from an LM_u by the target user as "information_u" (= i_u). A parallel distinction applies to LM_l s and LM_u s.

4. Types of information, informational dimensions and types of lexicographic messages

4.1 Types of information

When information as a phenomenon in the TLC is considered, a distinction between *information categories* (also referred to as *information types*) and *informational dimensions* is necessary.

Lexicographical information can be categorised into types like information about parts of speech of lexical items, morphological information about plural formations of lexical items, information about the meaning of a lexical item, etc.

The just mentioned types of information have in common that they can all be regarded as types of factual information, as opposed to, for example, types of affective information. They therefore belong to one and the same informational dimension.

4.2 Informational dimensions and types of lexicographic messages

Janssen (1996) and Steehouder et al. (1999) describe the elements of communication in the business or professional situation, as opposed to communication in informal and intimate situations on the one hand and mass communication on the other hand. In so doing, Steehouder et al. (1999: 24) introduces four *types of messages in an utterance*. Janssen introduces identical types, but at the information stage of the communication sequence. To distinguish these from the (subordinate) information types as described in 4.1 above, they will be referred to in the TLC as *informational dimensions*. The types that Janssen (1996) and Steehouder et al. (1999) introduce are adapted below for the purposes of lexicographical communication. The following four dimensions will be discussed:

- Referential information (4.2.1): the factual information encoded in an LM;
- Expressive information (4.2.2): information about the lexicographer as sender;
- Relational information (4.2.3): information about how the lexicographer regards the target user;
- Appeal information (4.2.4): information about the purpose(s) of an LU.

4.2.1 Informational dimension: referential information

Referential information can be described as the factual information encoded in a lexicographic message. (This seems to be what at least the function theory refers to as *data*, but *data* in the function theory are in fact LUs in the TLC.⁶) Consider a dictionary article with the lemma **vuvuzela**. The lemma sign is LU₄, in which at least two lexicographic messages, LM_l-1 and LM_l-2, are encoded. In both LM_ls referential information_l (= ref._l) is constructed and therefore both LM_ls are of the type *referential* LM_l (= ref.LM_l). For analytical purposes the following formal representation can be offered⁷:

- (17) LU₄ = lemma sign: "vuvuzela"
ref.LM_l-1[LU₄]: This dictionary article contains information about the lexical item *vuvuzela*.
ref._l-1[ref.LM_l-1[LU₄]] = This dictionary article contains information about the lexical item *vuvuzela*.
ref.LM_l-2[LU₄] = The lexical item *vuvuzela* is spelt v-u-v-u-z-e-l-a.

ref. i_l -2[ref.LM l -2[LU 4]] = The lexical item *vuvuzela* is spelt v-u-v-u-z-e-l-a.

In following the definition of *message* in (5) and therefore also the principle of expressibility as applied to the TLC, the value of i_l - x is identical to that of LM l - x , since LM l - x is the expression of i_l - x . The statement of the value of i_l - x is therefore redundant in further formal representations, and is henceforth omitted; the value is represented by LM l - x .

The value of each ref.LM l can be evaluated qualitatively and quantitatively in terms of referential criteria like accuracy and completeness. Such evaluations apply by extension to the relevant LUs as the observable manifestations of the messages, for which Grice's maxims in (3) could be adopted. In semantic terms the ref.LM l s could be referred to as the propositions of an LU (cf. e.g. Lyons 1977, Cruse 2011), and in terms of speech act theory they could be regarded as locutionary speech acts (cf. e.g. Levinson 1983, Cruse 2011).

4.2.2 Informational dimension: expressive information

A (lexicographic) utterance not only conveys referential information, but also information about the sender/lexicographer. This information belongs to the informational dimension of *expressive information* (= exp.i), and a unit of expressive information is constructed in an *expressive LM* (= exp.LM). By means of the LUs in a dictionary, a particular *identity* of the lexicographer (and/or publisher) is conveyed to the user. In business communication a distinction is made between a *corporate identity* and a *corporate image*. Fielding and Du Plooy-Cilliers (2014: 285) describe corporate identity as "everything that is distinctive, lasting and central to an organisation that gives it a particular character" and corporate image as "how the public views the organisation." Applying this distinction to lexicographical communication, the exp.LMs that the lexicographer encodes in LUs can be regarded as expressions of the lexicographer's *identity*, while the exp.LMs that the user decodes from LUs collectively form the *image* of the lexicographer with the user.

In the development of a dictionary plan and style guide, the lexicographer and publisher should determine what type of identity they would like to convey about themselves to the target user and formulate it in a *set of expressive style guide policy statements* (= exp.SPS{n}). (Identity is conveyed at various textual and structural levels in the dictionary.) This facet is perhaps less obvious in document types that Carstens and Van de Poel (2010: 68) refer to as *informative referential texts* (among which they include dictionaries), since the genres belonging to this type of text are seemingly designed to convey referential information and not to express any particular identity of the sender. However, the text type does not preclude the fact that expressive information is also conveyed, even if the sender is not aware of this fact or has not established an identity that he/she would like to convey.⁸ Compare Moon (2014) for a detailed

discussion of this type of information in learners' dictionaries from the perspective of critical lexicography. (Publishers are generally acutely aware of this informational dimension, especially in planning the marketing of their dictionaries.⁹)

According to Steehouder et al. (1999), receivers are sensitive to the expressive message(s) in an utterance and are often involuntarily influenced by them. Both the sender's and the receiver's *frame of reference* (= FoR) come into play here, since messages are constructed and encoded in utterances against the sender's FoR, and utterances are decoded and messages reconstructed against the receiver's FoR. This point will become pertinent below.

Janssen (1996) and Steehouder et al. (1999) identify the following characteristics of senders, among others, that can be regarded as expressive information that contribute to senders' identities and images: expert, novice, serious, self-assured, unsure, open, closed, arrogant, humble, formal, informal, long-winded, curt, grumpy, sympathetic, helpful, competent, incompetent, objective, subjective, credible.¹⁰ From a lexicographical perspective characteristics like prescriptive, descriptive and proscriptive (cf. Bergenholtz 2003) could be added. The list is not exhaustive.

Expressive information also conveys the sender's norms and values (as parts of the sender's identity), which may or may not be representative of the relevant speech community's or the target user's social and personal norms and values (cf. also Moon 2014). To relate this informational dimension to lexicographical communication, compare Van Sterkenburg (2003: 8), who comments in reference to the general monolingual dictionary that

[i]t also serves as a guardian of the purity of the language, of language standards and of moral and ideological values because it [i.e. the lexicographer – HLB] makes choices, for instance in the words that are to be described.

When exp.i_u is identical to the norms and values of the user, or it is the exp.i that the user expects to derive from lexicographical communication, usually no observable effect would be registered with the user. Effects are more readily observed when exp.i_u does not conform with the user's norms and values or is not the exp.i that the user expects to derive. For this reason the scope of effects in figure 1 is indicated by a dotted line. This aspect is central to the hypotheses in (27) to (29) below and in the discussion of unintentional effects in section 5.

Generally, professional lexicographer l wants to communicate at least the following exp.LM_l s to the user, which follow from the relevant $\text{SPS}\{n\}$:

- (18) a. $(\text{SPS-1} \rightarrow \text{exp.LM}_l\text{-1}) = l$ has zero bias.
b. $(\text{SPS-2} \rightarrow \text{exp.LM}_l\text{-2}) = l$ is objective.

These exp.LM_l s, among others, are assumed here to be identical to the exp.LM_u s that users generally expect to decode from LUs in professional dictionaries.¹¹ They may therefore be regarded as default exp.LM_u s: The user will only realise their value if exp.LM_u s with different or negating values replace them.

Consider the following lexical paraphrase of the lemma **vuvuzela**, taken from the online *Collins English Dictionary*:

- (19) LU₅ = lexical paraphrase: "an elongated plastic instrument that football fans blow to make a loud noise similar to the trumpeting of an elephant"

LU₅ could be partially decoded as follows:

- (20) ref.LM_u-1[LU₅] = A vuvuzela is an instrument.
ref.LM_u-2[LU₅] = A vuvuzela is made of plastic.
ref.LM_u-3[LU₅] = A vuvuzela has an elongated form.
...
ref.LM_u-n[LU₅] = The loud noise of a vuvuzela is similar to the trumpeting of an elephant.
exp.LM_l-1[LU₅] = *l* is not biased.
exp.LM_l-2[LU₅] = *l* is objective.

In this case

- (21) a. exp.LM_l-1[LU₅] = exp.LM_u-1[LU₅]
b. exp.LM_l-2[LU₅] = exp.LM_u-2[LU₅]

Consider, now, the following lexical paraphrase, one of several addressed at the lemma **vuvuzela** in the crowd-sourced online *Urban Dictionary*:

- (22) LU₆ = lexical paraphrase: "A mind-numbing torture device made of cheap, brightly colored [sic] plastic. It resembles a horn but its pitch cannot be changed. It is being used during the 2010 World Cup in South Africa."

LU₆ could be partially decoded as follows:

- (23) ref.LM_u-1[LU₆] = A vuvuzela is a mind-numbing torture device.
ref.LM_u-2[LU₆] = A vuvuzela is made of cheap, brightly coloured plastic.
ref.LM_u-3[LU₆] = A vuvuzela resembles a horn.
ref.LM_u-4[LU₆] = A vuvuzela's pitch cannot be changed.
ref.LM_u-5[LU₆] = A vuvuzela is being used during the 2010 World Cup in South Africa.
exp.LM_l-1[LU₆] = *l* is biased.
exp.LM_l-2[LU₆] = *l* is subjective.
exp.LM_l-3[LU₆] = *l* dislikes vuvuzelas.
exp.LM_l-4[LU₆] = *l* is offensive.
exp.LM_l-5[LU₆] = *l* is unfriendly.
...

Kosch (2013) raises the notion of *user expectation* in dictionaries, which can either be adapted for the TLC or utilised in its current conceptualisation. It is

important to recognise that any LM_{u-x} that is decoded from an LU_y follows from the user's frame of reference (= FoR_u), which contains the user's relevant norms, values, attitudes, etc.¹² This important relation can be formalised as follows:

$$(24) \quad FoR_u \rightarrow LM_{u-x}[LU_y]$$

The value of $LM_{u-x}[LU_y]$ is then measured against the user's *set of user expectations* (= $Xpt_u\{n\}$). If these expectations are fulfilled, non-observable effects (or zero-effects) can be predicted because the users would generally not display a response when their $Xpt_u\{n\}$ s are fulfilled. However, if they are not fulfilled, certain unpredictable and observable effects will occur.

In the case of $exp.LM_{u-1..5}[LU_6]$ in (23), it seems that SPS-1,2 do not apply to *Urban Dictionary*, as no gatekeeping appears to be in operation, judging by the contents of some of the lexical paraphrases.¹³ Therefore, the normal subset of *expressive user expectations* (= $exp.Xpt_u\{n\}$) is not fulfilled, which leads to a number of unpredictable effects (= $exp.E_{u-1..n}$) that follow from $exp.LM_{u-s}$. This can be represented in the following formalisation:

$$(25) \quad \begin{aligned} &\text{If } exp.LM_{u-1..5}[LU_6 \in \textit{Urban Dictionary}] \rightarrow exp.E_{u-1..n} \\ &\text{and } exp.LM_{u-1..5}[LU_6 \in \textit{Urban Dictionary}] < exp.Xpt_u\{n\} \\ &\text{then } exp.E_{u-1..n} = \text{unpredictable } exp.E_{u-1..n} \end{aligned}$$

("x \in y" reads "x is/as an element of (set) y"; "x < y" reads "x does not fulfil the requirements of (is less than) y"; "x{n}" reads "a set that contains elements of the type x".)

While:

$$(26) \quad \begin{aligned} &\text{If } exp.LM_{u-x}[LU_x \in \textit{professional dictionary } D] \rightarrow exp.E_{u-1..n} \\ &\text{and } exp.LM_{u-x}[LU_x \in \textit{professional dictionary } D] \cong exp.Xpt_u\{n\} \\ &\text{then } exp.E_{u-1..n} = \emptyset \end{aligned}$$

("x \cong y" reads "x fulfils the requirements of (is congruent to) y"; " \emptyset " reads "zero (i.e. non-observable)".)

From these expressions and the preceding discussion, the TLC could generate the following hypotheses:¹⁴

$$(27) \quad \begin{aligned} H_1: \quad &\text{If } exp.LM_{u-x}[LU_z] \rightarrow exp.E_{u-1..n} \\ &\text{and } exp.LM_{u-x}[LU_z] \cong exp.Xpt_u\{n\} \\ &\text{then } exp.E_{u-1..n} = \emptyset \end{aligned}$$

Conversely:

$$(28) \quad \begin{aligned} H_2: \quad &\text{If } exp.LM_{u-x}[LU_z] \rightarrow exp.E_{u-1..n} \\ &\text{and } exp.E_{u-1..n} = \emptyset \\ &\text{then } exp.LM_{u-x}[LU_z] \cong exp.Xpt_u\{n\} \end{aligned}$$

And:

- (29) H₃: If $\text{exp.LM}_{u-x}[\text{LU}_z] \rightarrow \text{exp.E}_{u-1..n}$
and $\text{exp.E}_{u-1..n} \neq \emptyset$
then $\text{exp.LM}_{u-x}[\text{LU}_z] < \text{exp.Xpt}_u\{n\}$

H₂ and H₃ can be expressed in natural language as follows:

- (30) H₂: If certain expressive user effects follow from a certain expressive LM decoded from a certain LU, and these expressive user effects are zero-effects, it follows that the expressive LM fulfils the set of expressive user expectations.
- H₃: If certain expressive user effects follow from a certain expressive LM decoded from a certain LU, and these expressive user effects are non-zero-effects, it follows that the expressive LM does not fulfil the set of expressive user expectations.

Effects (and zero-effects) are addressed in section 5.

The $\text{ref.LM}_{u-1..5}[\text{LU}_6]$ have not been addressed, but for the purposes of the current discussion they can be regarded as irrelevant, because (a) this section deals with exp.LMs only, and (b) they are probably of lesser importance when the purpose(s) of LU_6 are considered (cf. 4.2.4), although the truth of the propositions could be questioned.

The analyses of $\text{LU}_{5,6}$ in (20) and (23) above have demonstrated that the lexicographer does encode something of him-/herself in LUs (and obviously deliberately so in the case of LU_6). It is important to recognise this aspect in professional lexicography and consequently to devote the necessary time and energy during the development of the dictionary plan and style guide to formulate appropriate $\text{exp.SPS}\{n\}$ pertaining to the desired identity of the lexicographer, the dictionary and the publisher. Even in the comparatively simple LU_3 "n." is encoded the default $\text{exp.LM}_l[\text{LU}_3] = "l \text{ is an authority on the treated lexicon."}$

The propositions of exp.LMs can be evaluated qualitatively and quantitatively in terms of appropriateness for the user situation. In assigning values during formal evaluation, the evaluator should be able to validate these values in terms of pre-established criteria formulated against the background of the purpose(s) of the relevant dictionary, the target user and the user situation.

4.2.3 Informational dimension: relational information

In addition to information about the sender that is derivable from his/her utterances, Steehouder et al. (1999) and Janssen (1996) maintain that information about how the sender regards the receiver can also be derived from an utterance. This is referred to as *relational information* (= rel.i). In lexicographical

communication, LMs in or from which rel.i is constructed or reconstructed are accordingly called *relational* LMs (= rel.LMs). Relational messages can be overtly but also very subtly (and unintentionally) encoded in utterances. Like all types of messages, relational messages are decoded by receivers against their FoRs. Any rel.LM_u results in one or more *relational user effects* (= rel.E_u-1..n):

$$(31) \text{ rel.LM}_{u-x} \rightarrow \text{rel.E}_{u-1..n}$$

Steehouder et al. (1999) names four types of suppositions about the receiver that can influence communication. These are suppositions about

- the receiver's knowledge;
- the receiver's language skills;
- the receiver's norms, values, preferences and tastes; and
- the relation between the sender and the receiver.

Janssen (1996) identifies the following characteristics of the receiver as assigned by the sender that can be regarded as instances of rel.i following from the above suppositions: equal, superior, subordinate, rational, naïve, cooperative, stubborn, friend, enemy, intelligent, stupid. The list is not exhaustive.

As far as lexicographical communication is concerned, at least the first three types of suppositions above relate to the sociology of the dictionary user (cf. Wiegand 1977, Hartmann 1989). The relevant values should be based on empirical user research and be represented as elements of the target user profile (in terms of rel.SPS{n}) when a dictionary is planned, and should not be left to supposition (cf. e.g. Tarp (2008)).

As in the case of exp.LMs, rel.LMs are encoded in and decoded from LUs against the FoRs of the lexicographer and user respectively. Given the requirement of a target user profile in the planning of a dictionary, the lexicographer should be aware of the target user's FoR. Nevertheless, the lexicographer can (unintentionally) encode rel.LM_l-1..n in LUs based upon suppositions about the perceived relation between the lexicographer and the sender which may not follow from a rel.SPS{n} and do not fulfil the target user's set of relational expectations (= rel.Xpt_u{n}). It can also be the case that in following the dictionary plan or style guide, the lexicographer (unintentionally) encodes a rel.LM_l in an LU that does not fulfil the target user's rel.Xpt{n}. Obviously, it may also be the case that the user decodes rel.LMs from LUs that the lexicographer did not intend to encode in them. The question is then whether FoR_l = FoR_u with regard to the relevant elements in the relevant instance.

Louw and Beyer (in preparation) report that during empirical user research for a school dictionary

- (32) some of the respondents were of the opinion that a dictionary containing FSDs [full-sentence definitions – HLB] would be a "dumbed down" version of a "real" dictionary and would in effect "talk down" to the user.

The respondents were not the primary target users (i.e. learners), but a number of their teachers. In so far as their responses could be regarded as representative of the primary target user group¹⁵ and therefore as observable *relational user effects* (= rel.E_u-1..n), at least the following rel.LM_us can be derived as decoded by some of them from any particular full-sentence lexical paraphrase (LU_x):

- (33) rel.LM_u-1[LU_x] = *u* is stupid.
 rel.LM_u-2[LU_x] = *u* is not capable of using a "real" dictionary.
 rel.LM_u-3[LU_x] = *u* is socially inferior.

(While exp.LMs are expressed as propositions with *l* as the subject of the propositional sentence, rel.LMs are expressed with *u* as the subject, i.e. as propositions representing the lexicographer's view of the target user.)

It is to be expected that the rel.LMs encoded by the lexicographer in any LU are elements of the relevant rel.SPS{n}, and that rel.LM_u-1, 2, 3 above as decoded by the user from that LU are not elements of this rel.SPS{n}. Therefore, the relational LMs encoded by the lexicographer in the LU are not equal to rel.LM_u-1, 2, 3 above as decoded by the user from the LU. rel.E_u-1..n follow from rel.LM_u-1, 2, 3 above as decoded from the LU by the user, and a number of *intentional relational effects* (= rel.E_l-1..n) ideally follow from the rel.LMs encoded by the lexicographer in the LU. Therefore, the rel.E_u-1..n that could be observed with the users are not elements of rel.E_l{n} intended by the lexicographer, as elements of the rel.SPS{n}. (In non-technical terms: The lexicographer did not intend for rel.LM_u-1, 2, 3 above to be decoded from LU_x.) This explanation can be formalised in the following propositional formulas:

- (34) rel.LM_l-1..n[LU_x] ∈ rel.SPS{n}
 rel.LM_u-1,2,3[LU_x] ∉ rel.SPS{n}
 ∴ rel.LM_l-1..n[LU_x] ≠ rel.LM_u-1,2,3[LU_x]
 If rel.LM_u-1,2,3[LU_x] → rel.E_u-1..n
 and rel.LM_l-1..n[LU_x] → rel.E_l-1..n
 rel.E_u-1..n ≠ rel.E_l-1..n

Still, rel.LM_u-1, 2, 3 were decoded, which resulted in rel.E_u-1..n reported informally in (32). This case illustrates that, even with the best (educational) intentions, the rel.LM_ls encoded in an LU are not necessarily equal to the rel.LM_us decoded from the same LU. This demonstrates the necessity of establishing the target user's FoR empirically. Following this, the dictionary plan and/or style guide should formulate a rel.SPS{n} relating to the desirable rel.LM_ls to be encoded in LUs, as was alluded to earlier.

Hypotheses similar to those formulated with regard to exp.LMs in (27) to (29) in section 4.2.2 can be formulated for rel.LMs and the set of relational user expectations (= rel.Xpt_u{n}), but they will not be offered here as it is assumed that those in section 4.2.2 sufficiently demonstrate the relevant potential.

As is the case with exp.LMs, the values of rel.LMs can be evaluated qualitatively and quantitatively in terms of appropriateness for the user situation, following from FoR_u. In assigning values during formal evaluation, the evaluator should be able to validate these values in terms of pre-established criteria formulated against the background of the purposes of the relevant dictionary, the target user and the user situation.

At this point it should be recognised that exp.SPS{n} and rel.SPS{n} pertaining to a particular dictionary *D* constitute the identity of the lexicographer/publisher (= ID.lex_l) to be encoded in the dictionary (through its LUs), while the exp.LM_u-1..n and rel.LM_u-1..n decoded from the set of LUs in dictionary *D* by the user contribute to the image (= IM.lex_u) of the lexicographer/publisher pertaining to that dictionary:

$$(35) \quad \begin{aligned} \text{a. } & (\text{ID.lex}_l \gg D) = (\text{exp.SPS}\{n\} \gg D) \cup (\text{rel.SPS}\{n\} \gg D) \\ \text{b. } & (\text{IM.lex}_u \gg D) = (\text{exp.LM}_u\{n\}[\text{LU}\{n\} \in D]) \cup (\text{rel.LM}_u\{n\}[\text{LU}\{n\} \in D]) \end{aligned}$$

("x » y" reads "x, which is associated with y"; x{n} ∪ y{n} reads "the union of set x and set y".)

The image of the target user that is held by the lexicographer (= IM.usr_l) is formed by rel.SPS{n} pertaining to the particular dictionary *D*:

$$(36) \quad (\text{IM.usr}_l \gg D) = (\text{rel.SPS}\{n\} \gg D)$$

4.2.4 Informational dimension: appeal/purpose information

It is generally agreed that communication takes place with some purpose in mind (cf. Steinberg 2007). Appeal information is information about the purpose(s) of an utterance, and in lexicographical communication this type of information is constructed in *appeal* LMs (= aim.LM-1..n).

Steehouder, et al. (1999) distinguishes five types of communicative purposes of utterances, as follows:

Table 1: Types of communicative purposes according to Steehouder et al. (1999: 26)

Type of purpose	Purpose of the type of purpose
Informative	To inform the receiver about facts.
Instructive	To equip the receiver with a particular skill or set of skills.
Persuasive	To influence the receiver's opinion or attitude about something.
Motivational	To influence the receiver's behavioural intension, willingness.
Affective	To influence the receiver's emotions.

This typology of communicative purposes seems to be expressed in terms of the effect that a particular instance of communication is meant to have on the receiver. Given that Steehouder et al. focuses on professional communication, it is a valid typology, since the purpose of an utterance is not always clear in these types of situations, or in any typical interpersonal communication situation for that matter. In many cases the receiver has to identify the underlying communicative purpose of an utterance in order to give it meaning.

For lexicography, however, discrete classes of text segments (i.e. LUs in terms of the TLC) with aligned purposes have been identified and described; compare, for example, the more than 400 classes in the lexicographical text theory of Wiegand (1990, 1996, etc.) that are named according to their so-called genuine purpose. The text theory would refer to the class of LU₃ above ("n.") as *item giving the word class* (Wiegand 1996: 149). The statement of the genuine purpose of a class of text segments therefore refers to what type of referential message is encoded therein, and not what the communicative purpose of the LU would be in terms of intended effects. In some cases the class name includes reference to specific additional information (i.e. except what the text segment would propositionally state) that the user would be able to derive from the relevant text segment; compare the class name *item giving the definite article, which allows to ascertain the gender and the word class* [sic] (Wiegand 1996: 148; my underlining – HLB).

In section 4.2.1 above it was stated that ref.LM_I-1..n could in terms of the speech act theory in pragmatics be regarded as locutionary speech acts. Levinson (1983: 236) distinguishes as follows between the three types of acts that are performed with a linguistic utterance:

- (37) a. **locutionary act:** the utterance of a sentence with determinate sense and reference
- b. **illocutionary act:** the making of a statement, offer, promise, etc. in uttering a sentence, by virtue of the conventional *force* associated with it (or with its explicit performative paraphrase)
- c. **perlocutionary act:** the bringing about of effects on the audience by means of uttering the sentence, such effects being special to the circumstances of utterance

Illocutionary and perlocutionary acts could also be applied to lexicographical communication in the TLC. The so-called genuine purpose of a text segment (as expressed in its class name) in the text theory could be regarded as the illocutionary act of the relevant LU in speech act theory, while each communicative purpose of Steehouder, et al. in table 1 above could be regarded as a possible perlocutionary act of the relevant LU. Each of these speech acts can be assigned equivalents in the TLC, as indicated in the speech act theory analysis of LU₃ in (10) in table 2:

Table 2: Applying basic speech act theory (SAT) with TLC equivalents to an analysis of LU₃

LU ₃ \LU ₁ = item giving the word class: "n."		
SAT: Locutionary act	SAT: Illocutionary act	SAT: Perlocutionary act
TLC: Propositional contents	TLC: Purpose	TLC: Intentional effect
The word class of lexical item <i>book</i> is noun.	Lexicographer <i>l</i> STATES that the word class of lexical item <i>book</i> is noun.	User <i>u</i> is INFORMED of the fact that the word class of lexical item <i>book</i> is noun.

LU₃ would therefore be analysed as follows in the TLC:

- (38) LU₃\LU₁ = item giving the word class: "n."
 ref.LM_{*l*-1}[LU₃\LU₁] = The word class of lexical item *book* is noun.
 exp.LM_{*l*}{n}[LU₃\LU₁] ∈ exp.SPS{n}
 rel.LM_{*l*}{n}[LU₃\LU₁] ∈ rel.SPS{n}
 aim.LM_{*l*-1}[LU₃\LU₁] = **┐**(ref.LM_{*l*-1}[LU₃\LU₁])

The notation convention used to express the value of aim.LM_{*l*-1}[LU₃\LU₁] follows Searle (1969: 31): "**┐**" reads "STATEMENT". aim.LM_{*l*-1}[LU₃\LU₁] therefore reads: "The first (and only) purpose of LU₃, which is addressed at LU₁, is to STATE that ref.LM_{*l*-1}[LU₃\LU₁]." (The intentional effect is not included in this analysis – cf. section 5 below.) If an LU's purpose is to STATE, such statement does not pertain exclusively to propositions formulated in ref.LM_{*l*-1}.n, for if this were the case, a distinction between ref.LMs and aim.LMs would be baseless. Applying speech act theory to linguistic communication, the utterance "It is cold in here" consists of the locutionary act (propositional content) "It is cold in here," while the illocutionary act could be "STATEMENT(It is cold in here)" or "REQUEST(*Hearer* closes the door)", among others. The valid illocutionary act would be determined by the discourse situation. In the same way a specific LU can also serve purposes other than merely stating its propositional contents. Each purpose would be encoded in a separate aim.LM_{*l*}. Consider the text theory's *item giving the definite article, which allows to ascertain the gender and the word class* [sic] in the following dictionary article of a hypothetical German-English dictionary *GED*:

- (39) **Buch**, *das* book

The dictionary article consists of four categorised LUs:

- (40) a. LU₇ = item giving the form of the lemma sign: "Buch"
 b. LU₈ = non-typographical structural marker: ", "
 c. LU₉\LU₇ = item giving the definite article: "das"
 d. LU₁₀\LU₇ = item giving the lexical translation equivalent: "book"

Assuming that the default set of propositions representing $ID.lex_l$ is true by virtue of omitting $exp.LM_l\{n\}$ and $rel.LM_l\{n\}$, LU_9 can be analysed as follows:

- (41) $LU_9 \setminus LU_7$ = item giving the definite article: "das"
 $ref.LM_l-1[LU_9 \setminus LU_7]$ = Lexical item *Buch* takes the definite article *das*.
 $aim.LM_l-1[LU_9 \setminus LU_7]$ = $\vdash(ref.LM_l-1(LU_9 \setminus LU_7))$
 $aim.LM_l-2[LU_9 \setminus LU_7]$ = $\vdash(\text{The word class of lexical item } Buch \text{ is noun.})$
 $aim.LM_l-3[LU_9 \setminus LU_7]$ = $\vdash(\text{The gender of lexical item } Buch \text{ is neuter.})$

An $aim.LM_l$ offers an answer to a *potential user situation question* (= $p.SQ_u$) that expresses a particular potential information need of a target user in the typical user situation (following the principle of expressibility). In the ideal instance of lexicographical communication $aim.LM_u-y$ (decoded from the relevant LU by the user) has the same value as $aim.LU_l-x$ (encoded in the same LU by the lexicographer). Applying the formalisation in (41) to *GED*, an example could be:

- (42) $p.SQ_u \setminus GED \rightarrow aim.LM_u-1[LU_9 \in GED]$
 and $p.SQ_u \setminus GED = ?(\text{The word class of lexical item } Buch \text{ is } \dots)^{16}$
 $aim.LM_u-1[LU_9 \in GED] = aim.LM_l-2[LU_9 \setminus LU_7 \in GED]$

("p.SQ_u \setminus GED" reads "p.SQ_u directed at dictionary GED"; "? (p)" reads "QUESTION (propositional function p)" – cf. Searle (1969: 31).)¹⁷

The illocutionary force STATEMENT is not the only one that could apply to the purposes of LUs. Examples of at least two other types are:

- (43) a. WARNING, e.g. when a stylistic-functional label¹⁸ like *offensive* is offered for the text production situation;
 b. RECOMMENDATION, e.g. when the proscriptive approach is applied in a comment on a certain form in an inserted inner text for the text production situation.

Owing to limited space, the types of purposes (i.e. illocutions) will not be elaborated on here.

5. Effects of lexicographical communication and lexicographic functions

The terms *effect* and *intentional effect* have been used intuitively in earlier sections. These terms will now be clarified.

As the definition of the concept *communication* in (1) and the lexicographical communication model in figure 1 indicates, the sender communicates with the receiver in order to cause a particular *effect* with the receiver. Steinberg (2007: 21-

22) distinguishes between two types of effects, i.e. *intentional effects* and *unintentional effects*, which are defined as follows:

- (44) a. **Intentional effects** are the direct and predictable changes in the behaviour, opinion, attitudes or feelings of people in response to communication messages.
- b. **Unintentional effects** are the indirect influences and unpredictable results of the communication.

Whether an instance of communication has resulted in intentional and/or unintentional effects can be determined by the receiver's *feedback*, which Steinberg (2007: 50) describes as "the response of participants [in communication] to each other." In lexicographical communication, however, it is possible that an LU could have a particular effect on the target user without that effect being observable, as was suggested in 4.2.2 above.¹⁹ Effects of lexicographical communication could therefore be classified into two further subtypes, namely *observable effects* and *non-observable effects* (or *zero-effects*). The basic typology can be presented in figure 2 below:

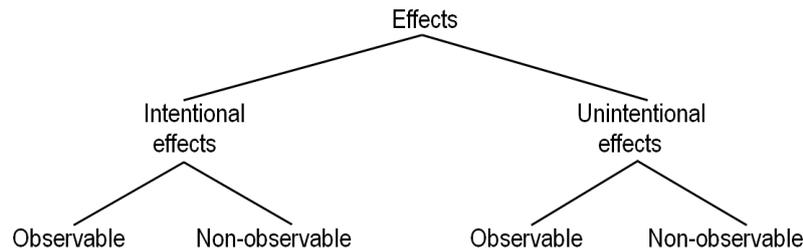


Figure 2: Basic typology of the effects of lexicographical communication

The type *non-observable intentional effect* might seem to be counter-intuitive, as the sender would normally like to observe some kind of effect in order to determine whether the communication effort has been successful. In lexicographical communication, however, a non-observable effect might in specific cases represent the intentional effect, which suggests that any observable effect would be an unintentional effect. This notion is evidenced in the hypotheses in (27) to (29) with regard to exp.LMs in 4.2.2 above. Considering the complexity of human behaviour, however, there is always the possibility that an unintentional effect could be non-observable. This might seem to challenge the hypotheses even before they could be tested. Nevertheless, citing the same human behaviour, it could be argued that the norm is known only by its exceptions. For example, if someone walked down the street completely naked (thereby violating the social norm that people wear clothes in public), there would predictably be an observable effect among (at least some of) the other

people in the street exactly because of this exception. By contrast, and all other things being equal, the phenomenon of someone walking down the street while fully clothed would not result in any observable effect exactly because it conforms to the relevant social norm. The argument and example are of course not unproblematic, but it does raise the notion of norms, and this can be aligned to the set of user expectations (= $X_{pt_u}\{n\}$) discussed in 4.2.2. $X_{pt_u}\{n\}$, as it pertains to $exp.LM$ and $rel.LM$, could be regarded in lexicographical communication as products of socio-communicative norms following from FoR_u s. It could therefore be reasonably assumed that if $exp.E_{u-1..n}$ and $rel.E_{u-1..n}$ are zero-effects, they result from conformation to the relevant socio-communicative norms and are therefore in fact intentional effects (cf. H_2 in (28)). To cater for the possibility of non-observable unintentional effects resulting from non-conformity to the relevant socio-communicative norms, it is imperative to conduct empirical research in controlled environments where observable effects that might otherwise be non-observable are elicited by means of accepted scientific methods. Thorough empirical research into target users' collective FoR and resulting $X_{pt_u}\{n\}$ could also substantiate the conclusion that if $exp.E_{u-1..n}$ and $rel.E_{u-1..n}$ are zero-effects, they are in fact intentional effects. In the same vein, if $ref.LM_l-1..n$ in an LU are not challenged by unintentional $ref.E_{u-1..n}$, it can be reasonably assumed that the relevant $ref.LM_l-1..n$ are accepted by the user and are therefore true, correct, accurate, relevant, etc.

At this point the relation between LM_u and E_u can be formalised:

- (45) a. $ref.LM_u-x \rightarrow ref.E_u-y$ (= referential user effects)
 b. $exp.LM_u-x \rightarrow exp.E_u-y$ (= expressive user effects)
 c. $rel.LM_u-x \rightarrow rel.E_u-y$ (= relational user effects)
 d. $aim.LM_u-x \rightarrow aim.E_u-y$ (= appeal user effects)

The aforementioned reasonable conclusions can also be formalised provisionally:

- (46) a. If $E_u-x = \emptyset$
 $E_u-x = \text{intentional } E_u-x$

This would seem to suggest that the ideal effect of lexicographical communication is a zero-effect. This is not true, though. It should be borne in mind that $ref.E_u$, $exp.E_u$ and $rel.E_u$ are not the primary E_u of lexicographical communication. They can be regarded as secondary effects or side-effect types, albeit effects of no lesser importance for successful communication. The E_l that lexicographical communication aims to yield, are $aim.E_l$. In terms of the discussion in section 4.2.4 above, the communicative purposes presented by Steehouder et al. (1999) in table 1 (and partially applied in table 2) could be regarded as the types of $aim.E_l$, thus:

- (47) As an intentional effect of lexicographical communication, user u
 a. is INFORMED about a certain fact (cf. table 2);

- b. is INSTRUCTED in a certain skill and therefore possesses that skill;
- c. is PERSUADED to adopt an opinion about or attitude towards something;
- d. is MOTIVATED to perform or not perform a certain action; and/or
- e. experiences a certain EMOTION.

To determine if an instance of lexicographical communication has been successful, the lexicographer needs to establish if SQ_u has been answered. This can only be done if the effect of the particular instance of lexicographical communication can be observed. The types of $aim.E_l$ identified in (47) should therefore be expanded to include a component that refers to an observable effect, since being informed/instructed/persuaded/motivated or experiencing a certain emotion is in itself not automatically observable. The value of this additional component will be particular to each type of LU. Assigning intentional effects to the $LM_l\{n\}$ of LU_9 in (41) will result in the following analysis in terms of the TLC (focusing on the communicative user situation):

- (48) $LU_9 \setminus LU_7$ = item giving the definite article: "das"
 $ref.LM_l-1[LU_9 \setminus LU_7]$ = Lexical item *Buch* takes the definite article *das* in the nominative case.
 $\rightarrow (ref.E_l-1 \gg ref.LM_l-1[LU_9 \setminus LU_7]) = \emptyset$
 $exp.LM_l-1..n[LU_9 \setminus LU_7] \in exp.SPS\{n\}$
 $\rightarrow (exp.E_l-1 \gg exp.LM_l-1..n[LU_9 \setminus LU_7]) = \emptyset$
 $rel.LM_l-1..n[LU_9 \setminus LU_7] \in rel.SPS\{n\}$
 $\rightarrow (rel.E_l-1 \gg rel.LM_l-1..n[LU_9 \setminus LU_7]) = \emptyset$
 $aim.LM_l-1[LU_9 \setminus LU_7] = \vdash(ref.LM_l-1[LU_9 \setminus LU_7])$
 $\rightarrow (aim.E_l-1 \gg aim.LM_l-1[LU_9 \setminus LU_7]) = INFORM:(u COMBINES: the definite article *das* & lexical item *Buch* in the nominative case.) $\in tp.s_u$
 $\rightarrow (aim.E_l-2 \gg aim.LM_l-1[LU_9 \setminus LU_7]) = INFORM:(u APPLIES: the appropriate declension of the definite article *das* with lexical item *Buch* in the dative, accusative and genitive cases respectively) $\in tp.s_u$
 $\rightarrow (aim.E_l-3 \gg aim.LM_l-1[LU_9 \setminus LU_7]) = PERSUADE:(u EVALUATES: the combination of the definite article *das* with lexical item *Buch* in the nominative case = correct) $\in tr.s_u$
 $\rightarrow (aim.E_l-4 \gg aim.LM_l-1[LU_9 \setminus LU_7]) = PERSUADE:(u EVALUATES: the combination of a definite article \neg *das* with lexical item *Buch* in the nominative case = \neg correct) $\in tr.s_u$
 $\rightarrow (aim.E_l-5 \gg aim.LM_l-1[LU_9 \setminus LU_7]) = PERSUADE:(u EVALUATES: the application of the appropriate declension of definite article *das* with lexical item *Buch* in the dative,$$$$$

accusative and genitive cases respectively = correct) \in tr.s_u
 $\rightarrow (\text{aim.E}_{I-6} \gg \text{aim.LM}_{I-1}[\text{LU}_9 \setminus \text{LU}_7]) = \text{PERSUADE:}(u \text{ EVALUATES: the application of } \neg(\text{the appropriate declension of definite article } \textit{das}) \text{ with lexical item } \textit{Buch} \text{ in the dative, accusative and genitive cases respectively} = \neg\text{correct}) \in \text{tr.s}_u$
 $\text{aim.LM}_{I-2}[\text{LU}_9 \setminus \text{LU}_7] = \mathbf{\neg}$ (The word class of lexical item *Buch* is noun.)
 $\rightarrow (\text{aim.E}_{I-1} \gg \text{aim.LM}_{I-2}[\text{LU}_9 \setminus \text{LU}_7]) = \text{INFORM:}(u \text{ APPLIES: lexical item } \textit{Buch} = \text{the core of a noun phrase.}) \in \text{tp.s}_u$
 $\rightarrow (\text{aim.E}_{I-2} \gg \text{aim.LM}_{I-2}[\text{LU}_9 \setminus \text{LU}_7]) = \text{PERSUADE:}(u \text{ EVALUATES: the application (lexical item } \textit{Buch} = \text{the core of a noun phrase)} = \text{correct}) \in \text{tr.s}_u$
 $\rightarrow (\text{aim.E}_{I-3} \gg \text{aim.LM}_{I-2}[\text{LU}_9 \setminus \text{LU}_7]) = \text{PERSUADE:}(u \text{ EVALUATES: the application (lexical item } \textit{Buch} = \neg(\text{the core of a noun phrase})) = \neg\text{correct}) \in \text{tr.s}_u$
 $\text{aim.LM}_{I-3}[\text{LU}_9 \setminus \text{LU}_7] = \mathbf{\neg}$ (The gender of lexical item *Buch* is neuter.)
 $\rightarrow (\text{aim.E}_{I\{n\}} \gg \text{aim.LM}_{I-3}[\text{LU}_9 \setminus \text{LU}_7]) = (\text{aim.E}_{I\{n\}} \gg \text{aim.LM}_{I-1}[\text{LU}_9 \setminus \text{LU}_7])$

(" \neg " indicates negation.)

$\langle \text{aim.E}_{I-1} \gg \text{aim.LM}_{I-1}[\text{LU}_9 \setminus \text{LU}_7] \rangle^{20}$ above could be described as follows in natural language:

- (49) Intentional effect aim.E_{I-1} associated with LU_9 's purpose aim.LM_{I-1} is that the target user in the text production user situation (= tp.s_u) is INFORMED that the lexical item *Buch* takes the definite article *das* in the nominative case; therefore the target user COMBINES the definite article *das* with lexical item *Buch* in the nominative case.

In comparison $\langle \text{aim.E}_{I-3} \gg \text{aim.LM}_{I-1}[\text{LU}_9 \setminus \text{LU}_7] \rangle$ could be described as follows:

- (50) Intentional effect aim.E_{I-3} associated with LU_9 's purpose aim.LM_{I-1} is that the target user in the text reception user situation (= tp.s_u) is PERSUADED that the lexical item *Buch* takes the definite article *das* in the nominative case; therefore the target user EVALUATES as correct the combination of the definite article *das* with lexical item *Buch* in the nominative case.

(49) and (50) constitute two of the observable intentional effects of $\text{aim.LM}_{I-1}[\text{LU}_9 \setminus \text{LU}_7]$ that are diagnostic of successful lexicographical communication in the particular user situation. The key terms like *combine* and *evaluate* that form

part of the aim. E_l propositions relate directly to the observability and measurability of aim. E_l . They can be taken from education theory, for instance Bloom's revised taxonomy of educational objectives (cf. Anderson and Krathwohl 2001).²¹

It is now theoretically necessary to be able to express in a single statement the relation between a particular LU_x , a particular aim.LM $_{l-y}$ of LU_x and the aim. E_{l-z} that follows from that LU as associated with aim.LM $_{l-y}$. Such a statement represents the *lexicographic function* (= f) of LU_x and can be formalised as follows:²²

$$(51) \quad f[LU_x] = (\text{aim.LM}_{l-y}[LU_x] \rightarrow (E_{l-z} \gg \text{aim.LM}_{l-y}[LU_x]))$$

The function of LU_9 in (51) with regard to $\langle \text{aim.E}_{l-1} \gg \text{aim.LM}_{l-1}[LU_9 \setminus LU_7] \rangle$ can be expressed in natural language as follows:

- (52) The first function of the item giving the definite article *das* of the lemma **Buch** is to STATE that the lexical item *Buch* takes the definite article *das* in the nominative case so that the target user in the text production user situation is INFORMED that the lexical item *Buch* takes the definite article *das* in the nominative case; therefore, the target user COMBINES the definite article *das* with the lexical item *Buch* in the nominative case.

This particular function (= $f-x$) can be formalised as follows:

$$(53) \quad f-1[LU_9 \setminus LU_7] = (\text{aim.LM}_{l-1}[LU_9 \setminus LU_7] \rightarrow (\text{aim.E}_{l-1} \gg \text{aim.LM}_{l-1}[LU_9 \setminus LU_7]))$$

The various classification possibilities of functions will not be dealt with here.

For each LU a set of functions (= $f\{n\}$) can be formulated. Actual appeal user effects (aim. $E_{u-1..n}$) that result from a particular LU can then be compared against the sets of intended appeal effects (aim. $E_l\{n\}$) that constitute elements of the set of functions of the LU to determine whether lexicographical communication has been successful. The result of such a comparison will lead to the diagnosis of a particular aim. E_{u-x} as intentional or unintentional.

Consequently, an intentional actual appeal user effect can be referred to as a *functional appeal effect* ($f.\text{aim.E}_{u-y}$), which occurs when the actual appeal user effect following from a particular LU is an element of the set of intentional appeal effects of that LU (as an element of the set of functions of that LU):

$$(54) \quad \begin{aligned} &\text{If } LU_x \rightarrow \text{aim.E}_{u-y} \\ &\text{and } \text{aim.E}_{u-y} \in (\text{aim.E}_l\{n\}[LU_x] \in f\{n\}[LU_x]) \\ &\rightarrow \text{aim.E}_{u-y} = f.\text{aim.E}_{u-y} \\ &\therefore LU_x \rightarrow f.\text{aim.E}_{u-y} \end{aligned}$$

Unintentional actual user effects can be classified into two categories, i.e. *non-functional effects* and *dysfunctional effects*.

If actual effect aim. E_{u-y} follows from LU_x and is not an element of the set

of intended effects $\text{aim.E}_I\{n\}$ of LU_x (as an element of the set of functions of LU_x), then actual effect aim.E_{u-y} is a *non-functional* effect $\theta.\text{aim.E}_{u-y}$ that follows from LU_x :

$$(55) \quad \begin{aligned} &\text{If } \text{LU}_x \rightarrow \text{aim.E}_{u-y} \\ &\text{and } \text{aim.E}_{u-y} \notin (\text{aim.E}_I\{n\}[\text{LU}_x] \in f\{n\}[\text{LU}_x]) \\ &\rightarrow \text{aim.E}_{u-y} = \theta.\text{aim.E}_{u-y} \\ &\therefore \text{LU}_x \rightarrow \theta.\text{aim.E}_{u-y} \end{aligned}$$

If actual effect aim.E_{u-y} follows from LU_x and negates an intended effect that is an element of the set of intended effects $\text{aim.E}_I\{n\}$ of LU_x (as an element of the set of functions of LU_x), then actual effect aim.E_{u-y} is a *dysfunctional* effect $\Psi.\text{aim.E}_{u-y}$ that follows from LU_x :

$$(56) \quad \begin{aligned} &\text{If } \text{LU}_x \rightarrow \text{aim.E}_{u-y} \\ &\text{and } \text{aim.E}_{u-y} \rightarrow \neg(\text{aim.E}_I\{z\} \in (\text{aim.E}_I\{n\}[\text{LU}_x] \in f\{n\}[\text{LU}_x])) \\ &\rightarrow \text{aim.E}_{u-y} = \Psi.\text{aim.E}_{u-y} \\ &\therefore \text{LU}_x \rightarrow \Psi.\text{aim.E}_{u-y} \end{aligned}$$

Functional effects signal successful lexicographical communication and dysfunctional effects signal failed lexicographical communication. Non-functional effects are effects that were not foreseen within the framework of the functions of the relevant LUs, but that have not necessarily had an adverse effect on lexicographical communication. An example of a non-functional effect would be when a user uses a bilingual dictionary to find the meaning of a lexical item by deriving it from the translation equivalent provided (Beyer 2013). The function of the relevant translation equivalent is not to state the meaning of the lemma so that the user would be informed of its meaning to apply it in the user situation, but rather to state a translation equivalent so that the user would be informed of the translation equivalent and apply it in translating a text in the appropriate user situation. Although the user performed a successful dictionary consultation procedure and SQ_u was answered, it was not the "intention" of the lexicographer that this be the case; hence a non-functional effect occurred, even though the dictionary was useful to the user.²³

A non-functional effect could also turn out to be an effect that could be regarded as functional, but it was not predicted in formulating the particular set of functions. Such an effect could then be incorporated into that set of functions, changing its status from non-functional to functional.

To revisit ref.E_{u-} , exp.E_u and rel.E_u in terms of the paradigm of functionality, the following propositions apply:

$$(57) \quad \begin{aligned} &\text{a. If } \text{LU}_x \rightarrow \text{ref.E}_{u-y} \\ &\quad \text{and } \text{ref.E}_{u-y} = \emptyset \\ &\quad \rightarrow \text{ref.E}_{u-y} = f.\text{ref.E}_{u-y} \\ &\therefore \text{LU}_x \rightarrow f.\text{ref.E}_{u-y} \end{aligned}$$

- b. If $LU_x \rightarrow \text{exp.E}_{u-y}$
and $\text{exp.E}_{u-y} = \emptyset$
 $\rightarrow \text{exp.E}_{u-y} = f.\text{exp.E}_{u-y}$
 $\therefore LU_x \rightarrow f.\text{exp.E}_{u-y}$
- c. If $LU_x \rightarrow \text{rel.E}_{u-y}$
and $\text{rel.E}_{u-y} = \emptyset$
 $\rightarrow \text{rel.E}_{u-y} = f.\text{rel.E}_{u-y}$
 $\therefore LU_x \rightarrow f.\text{rel.E}_{u-y}$

When an unintentional (observable) effect occurs in any of these cases, that effect has to be analysed to arrive at the specific LM_{u-x} (and LU) from which the effect had followed. It can then be determined if the effect is non-functional or dysfunctional by measuring its value against the equivalent LM_{l-x} and the relevant $Xpt_u\{n\}$.

- (58)
- a. If $\text{ref.E}_u \neq \emptyset$
 $\text{ref.E}_u = \theta.\text{ref.E}_u \vee \Psi.\text{ref.E}_u$
 - b. If $\text{exp.E}_u \neq \emptyset$
 $\text{exp.E}_u = \theta.\text{exp.E}_u \vee \Psi.\text{exp.E}_u$
 - c. If $\text{rel.E}_u \neq \emptyset$
 $\text{rel.E}_u = \theta.\text{rel.E}_u \vee \Psi.\text{rel.E}_u$

("x \vee y" reads "x or y".)

6. The case of the vuvuzela

In this section the TLC will be applied to the research problem as a framework to analyse the criticism levelled at the example sentence *Vuvuzelas maak 'n groot lawaai by sokkerwedstryde* [\pm Vuvuzelas make a lot of noise at soccer matches] presented during the keynote address mentioned in the introduction, and to determine why the criticism did not seem to be validated or dealt with constructively. First, a partial formal TLC analysis of the example sentence will be provided, followed by a description of the criticism against the background of the prevailing context.

6.1 A formal TLC analysis of the relevant LU

The following partial formal analysis of the example sentence is presented in as far as the various LMs are assumed to be true, since the author of this article is not involved with the dictionary project in question.

(59) LU₁₀ = example sentence: "Vuvuzelas maak 'n groot lawaai by sokkerwedstryde"

ref.LM_l-1[LU₁₀] = Vuvuzelas make a lot of noise at soccer matches.

→ (ref.E_l-1»ref.LM_l-1[LU₁₀]) = ∅

exp.LM_l-1[LU₁₀] = *l* is an expert in the treated languages.

→ (exp.E_l-1»exp.LM_l-1[LU₁₀]) = ∅

exp.LM_l-2[LU₁₀] = *l* is objective.

→ (exp.E_l-1»exp.LM_l-2[LU₁₀]) = ∅

exp.LM_l-3[LU₁₀] = *l* has zero bias.

→ (exp.E_l-1»exp.LM_l-3[LU₁₀]) = ∅

exp.LM_l-4[LU₁₀] = *l* believes LU₁₀ to be true.

→ (exp.E_l-1»exp.LM_l-4[LU₁₀]) = ∅

rel.LM_l-1[LU₁₀] = *u* is a learner of either of the treated languages.

→ (rel.E_l-1»rel.LM_l-1[LU₁₀]) = ∅

rel.LM_l-2[LU₁₀] = *u* possesses the set of language skills LS_{*u*}{*n*} in either of the treated languages.

→ (rel.E_l-1»rel.LM_l-2[LU₁₀]) = ∅

rel.LM_l-3[LU₁₀] = *u* possesses the set of dictionary usage skills DS_{*u*}{*n*}.

→ (rel.E_l-1»rel.LM_l-3[LU₁₀]) = ∅

...

aim.LM_l-1[LU₁₀] = **⊢**(ref.LM_l-1[LU₁₀])

→ (aim.E_l-1»aim.LM_l-1[LU₁₀]) = INFORM:(*u* WRITES: LU₁₀.)

∈ tp.s_{*u*}

→ (aim.E_l-2»aim.LM_l-1[LU₁₀]) = PERSUADE:(*u* EVALUATES: LU₁₀ = grammatically correct/possible.) ∈ tr.s_{*u*}

aim.LM_l-2[LU₁₀] = **⊢**(The lexical item *vuvuzela* can fulfil the role of subject/agent in an Afrikaans sentence.)

→ (aim.E_l-1»aim.LM_l-2[LU₁₀]) = PERSUADE:(*u* WRITES: an Afrikaans sentence (≠ LU₁₀) with the lexical item *vuvuzela* in the role of subject/agent.) ∈ tp.s_{*u*}

→ (aim.E_l-2»aim.LM_l-2[LU₁₀]) = PERSUADE:(*u* EVALUATES: an Afrikaans sentence (≠ LU₁₀) with the lexical item *vuvuzela* in the role of subject/agent = grammatically correct) ∈ tr.s_{*u*}

aim.LM_l-3[LU₁₀] = **⊢**(The lexical item *vuvuzela* combines frequently with the verb *maak*.)

→ (aim.E_l-1»aim.LM_l-3[LU₁₀]) = PERSUADE:(*u* WRITES: an Afrikaans sentence (≠ LU₁₀) in which the lexical item *vuvuzela* combines with the verb *maak*.) ∈ tp.s_{*u*}

→ (aim.E_l-2»aim.LM_l-3[LU₁₀]) = PERSUADE:(*u* EVALUATES: an Afrikaans sentence (≠ LU₁₀) in which the lexical item *vuvuzela* combines with the verb *maak* = grammatically correct.) ∈ tr.s_{*u*}

...

According to this partial analysis LU₁₀ has at least 14 lexicographic functions, of which six are appeal or primary functions. The remaining eight may generally be regarded as default functions following from the dictionary's (assumed) ref.SPS{n}, exp.SPS{n} and rel.SPS{n}.

6.2 An informal representation of the actual effects

The criticism of the example sentence can be paraphrased as follows:

- (60) a. The delegate (as a potential user) expressed disappointment at the fact that he could not observe the notion of festivity and celebration, with which the vuvuzela is also associated, in the example. It seemed regrettable to him that the vuvuzela is associated only with excessive noise in the dictionary article.
- b. The word *noise* in the example sentence, instead of, for instance *sound*, seems to portray the vuvuzela in a negative light as making a "sound, esp one that is loud and disturbing" (*Collins English Dictionary*), "a loud or unpleasant sound" (*Macmillan*).

6.3 Context, frame of reference and user expectations

The interpretation of the effects of lexicographical communication should take place against the background of the prevailing context. In the context-change theory of speech acts, *context* is defined as "a set of propositions, describing the beliefs, knowledge, commitments and so on of the participants in a discourse" (Levinson 1983: 276). In applying this definition to lexicographical communication, it follows that context C is the union of the sets of propositions describing FoR_l and FoR_u respectively:

$$(61) \quad C = \text{FoR}_l\{n\} \cup \text{FoR}_u\{n\}$$

Context is anchored geographically, temporally and socially (i.e. culturally), since beliefs, knowledge, etc. vary geographically, temporally and socially. Ideally, FoR_l{n} = FoR_u{n}, but this is rarely the case in reality. Rather, lexicographical communication should strive towards the greatest possible intersection between FoR_l{n} and FoR_u{n} as informed by empirical research. This implies that there will be FoR elements that are not shared by the lexicographer and the target user, and that there might also be elements of FoR_l{n} that are not compatible with elements of FoR_u{n}. The following hypothetical FoR elements with regard to the vuvuzela as artefact are assigned to the lexicographer and the target user respectively for the purposes of this analysis:²⁴

- (62) a. FoR_l-1 = The vuvuzela is an instrument blown by fans at soccer matches.

- FoR_I-2 = Many people complain about the loud sound made by vuvuzelas.
- b. FoR_U-1 = The vuvuzela is an instrument blown by fans at soccer matches.
 FoR_U-2 = The vuvuzela is an icon of the 2010 Soccer World Cup tournament hosted by South Africa.
 FoR_U-3 = The vuvuzela is an icon of the atmosphere of celebration and festivity that prevailed during the 2010 Soccer World Cup tournament in South Africa.
 FoR_U-4 = The vuvuzela represents the global recognition of South Africa and Africa in 2010.
 FoR_U-5 = Many people complain about the loud sound made by vuvuzelas.

From the FoR_U{n} might have followed Xpt_U{n}:

- (63) FoR_U{n} → Xpt_U{n}
 Xpt_U{n} = {Xpt_U-1, Xpt_U-2}
 Xpt_U-1 = A South African dictionary treating the lexical item *vuvuzela* should refer to its iconic status. (»FoR_U-2, 3)
 Xpt_U-2 = A South African dictionary treating the lexical item *vuvuzela* should mention South Africa. (»FoR_U-4)

The intersection of FoR_I{n} and FoR_U{n} in (62) is limited to two elements:

- (64) FoR_I{n} ∩ FoR_U{n} = {(The vuvuzela is an instrument blown by fans at soccer matches.), (Many people complain about the loud sound made by vuvuzelas.)}

("x ∩ y" reads "the intersection of sets x and y".)

6.4 A formal analysis of the actual effects

The actual user effects in (60) can be represented formally as follows:

- (65) a. (E_U-1»LU₁₀) = (u EVALUATES: LU₁₀ = ¬accurate)
 b. (E_U-2»LU₁₀) = (u EVALUATES: prejudice ∈ LU₁₀)

Each of the actual user effects can be analysed formally in (66) and (67) respectively (formalisation in part a and natural language equivalent in part b):

- (66) a. (E_U-1»LU₁₀) = (u EVALUATES: LU₁₀ = ¬accurate)
 ∴ (E_U-1»LU₁₀) = ref.E_U-1²⁵
 ref.E_U-1 ≠ ∅
 ∴ ref.E_U-1 = θ.ref.E_U-1 ∨ Ψ.ref.E_U-1
 ref.LM_U-1 < ref.Xpt_U{n}

$$\begin{aligned} \text{ref.LM}_{u-1} &\rightarrow \text{ref.E}_{u-1} \\ \therefore \text{ref.E}_{u-1} &= \Psi.\text{ref.E}_{u-1} \\ \therefore (E_{u-1}\gg\text{LU}_{10}) &= (\Psi.E_{u-1}\gg\text{LU}_{10}) \end{aligned}$$

- b. Actual user effect $\langle E_{u-1}\gg\text{LU}_{10} \rangle$ comprises of the user evaluating LU_{10} as inaccurate. Based on the evaluative expression *inaccurate*, $\langle E_{u-1}\gg\text{LU}_{10} \rangle$ is a referential actual user effect ref.E_{u-1} . This effect is not a zero-effect; therefore it is an unintentional effect. Because ref.LM_{u-1} does not fulfil the set of referential user expectations $\text{ref.Xpt}_u\{n\}$, and ref.E_{u-1} follows from ref.LM_{u-1} , ref.E_{u-1} is a dysfunctional effect. Therefore, $\langle E_{u-1}\gg\text{LU}_{10} \rangle$ is a dysfunctional effect.

(67) a. $(E_{u-2}\gg\text{LU}_{10}) = (u \text{ EVALUATES: prejudice} \in \text{LU}_{10})$
 $\therefore (E_{u-2}\gg\text{LU}_{10}) = \text{exp.E}_{u-1}$
 $\text{exp.E}_{u-1} \neq \emptyset$
 $\therefore \text{exp.E}_{u-1} = \theta.\text{exp.E}_{u-1} \vee \Psi.\text{exp.E}_{u-1}$
 $\text{exp.LM}_{u-1} = \neg\text{exp.LM}_{l-3}[\text{LU}_{10}]$
 $\text{exp.LM}_{u-1} \rightarrow \text{exp.E}_{u-1}$
 $\therefore \text{exp.E}_{u-1} = \Psi.\text{exp.E}_{u-1}$
 $\therefore (E_{u-2}\gg\text{LU}_{10}) = (\Psi.E_{u-2}\gg\text{LU}_{10})$

- b. Actual user effect $\langle E_{u-2}\gg\text{LU}_{10} \rangle$ comprises of the user evaluating LU_{10} as containing prejudice. Based on the evaluative expression *prejudice*, $\langle E_{u-2}\gg\text{LU}_{10} \rangle$ is an expressive actual user effect exp.E_{u-1} . This effect is not a zero-effect; therefore it is an unintentional effect. Because exp.LM_{u-1} negates $\text{exp.LM}_{l-3}[\text{LU}_{10}]$, and exp.E_{u-1} follows from it, exp.E_{u-1} is a dysfunctional effect. Therefore, $\langle E_{u-2}\gg\text{LU}_{10} \rangle$ is a dysfunctional effect.

The analyses show that both actual user effects are dysfunctional effects, i.e. effects not intended by the lexicographer, but effects that have threatened successful lexicographical communication.

6.5 Why were the dysfunctional user effects not effectively neutralised?

When dysfunctional effects occur, they should ideally be neutralised. Effective neutralisation requires the lexicographer to conduct a thorough analysis of a dysfunctional effect, like in (66) and (67) above, to arrive at the precise cause.

In the case of the vuvuzela, the TLC analysis (and the time) was not available to effectively neutralise the dysfunctional effects. The lexicographer's defence was mounted from a referential perspective which included arguments that the example sentence was not in any way incorrect and was representative of an appropriate corpus analysis, but at least one of the dysfunctional effects resulted from expressive information. Criticism and defence were therefore at least partially misaligned and possible unrecognised differences between the

FoR_l and FoR_u hindered effective neutralisation further, since the $\chi_{pt_u}\{n\}$ was not known to the lexicographer.

6.6 Conclusion

It should be noted that the member of the conference audience who criticised the example sentence might not have been a member of the target user group, and that a typical target user might not have shared the delegate's FoR and $\chi_{pt_u}\{n\}$. It is therefore possible that the same dysfunctional effects might not have been observed with a target user.

As was stated in the introduction, however, the aim of this article is to present the TLC's capacity to scientifically address the case, and not to choose a side or to appoint a "winner". For this reason it is far less important that the stated LMs, FoRs and $\chi_{pt_u}s$ represent the actual LMs, FoRs and $\chi_{pt_u}s$ of the lexicographer and the delegate respectively than it is to demonstrate that the introduced theoretical framework facilitates scientific explanation. It would seem that exactly this has been demonstrated.

7. Applying the TLC to dictionary evaluation

The elements of the TLC introduced and applied in this article can be distilled to the following general questions that could guide formative and summative dictionary evaluation:

- (68)
- a. Do the LMs and especially the aim.LMs support the purposes of the dictionary?
 - b. Are the ref.LMs correct, accurate, true, etc.?
 - c. Do the exp.LMs express the appropriate and desired attitudes, norms and values of the lexicographer (and by extension that of society)?
 - d. Do the rel.LMs reflect the desired perceived relation between the lexicographer and the target user?

8. Perspective

This article represents the first introduction of the TLC as a possible alternative or at least a complement to existing lexicographical theories. Some elements of this developing theory have been applied tentatively and informally in previous publications as components of a so-called communicative metalexigraphy. The theoretical elements that are formally developed in this article are also successfully applied to a real lexicographical situation, which demonstrates its validity.

Currently, the TLC seems to be situated in the overlapping area between

semiotics and linguistics, which supports the argument that lexicography is a highly interdisciplinary field of study.

This article in particular has drawn significantly on professional communication and what could be referred to as basic classical speech act theory. Work ahead includes the full exploration of speech act theory and conversational implicature for the further development of the TLC, but also other theories in pragmatics, as well as the disciplines of text linguistics and document design. Furthermore, establishing the level of abstraction to which the relations between the sets of propositions and qualities developed in the process could be generalised would be central in an eventual pursuit of a more general theory of lexicographical communication.

The TLC is in its infancy, and as such offers exciting research potential. Nevertheless, this article has already demonstrated that this theory possesses explanatory power, and that is one of the key requirements for a scientific theory.

Notes

1. This seems to be the English translation of a work that had originally appeared in Chinese in 1991.
2. Beyer (in preparation) takes due cognisance of this work. For the purposes of this article an elaboration is not relevant here.
3. The principle of expressibility has been criticised by some scholars (cf. Levinson 1983). However, addressing the criticism falls outside the scope of this article.
4. Wiegand (e.g. 1990: 13-14) would regard the linguistic symbol "n." in LU₃ as an instance of the *textual condensation* of the linguistic symbols (i.e. the sentence) "The lexical item *book* is a noun" in LM_l, since LM_l is a full-sentence representation of LU₃. In the TLC, however, textual condensation is not recognised to the full extent that Wiegand describes it. In as far as "n." is an abbreviation of "noun", textual condensation does apply. However, a symbol like the tilde "~", traditionally used to represent the lemma sign, is not characterised as an instance of the textual condensation of the lemma sign, but rather as an instance of lexicographical anaphora (which can, like textual condensation, be employed to save space in a printed dictionary). The encoding of LM_l in LU₃ is preceded by an instance of what could be referred to as pre-encoding textual condensation: the symbol "noun" is textually condensed to the symbol "n." before the message "The lexical item *book* is a noun" is encoded in the symbol "n." to create LU₃. The encoding process is therefore not an instance of textual condensation (or of lexicographical anaphora). Rather, it becomes apparent that, apart from the code system (i.e. the language) that is described by means of lexicographical communication, the dictionary employs its own code system consisting of its own types of signs and symbols (e.g. "n." instead of "The lexical item *book* is a noun.") organised according to its own code (i.e. not exclusively the grammar of the language that is the subject of the dictionary, but also the "dictionary grammar" that is realised in the various lexicographical structures). In this way the TLC reveals that the dictionary user, in attempting to gain command of one code system (i.e. the treated language) by consulting the dictionary, first needs to sufficiently command the mediating hybrid code system (i.e. the lexicographical symbols and structures in conjunction with the relevant linguistic symbols and structures) that is employed in the

dictionary. This revelation could inform approaches to user guide design and dictionary pedagogies (that could, for example, be represented by slogans such as "Do you speak Lexicographish?"). This insight also facilitates the description of the problem with many user guides and dictionary pedagogies as comparable to a teacher starting the first lesson in English as a Foreign Language with the statement: $S \rightarrow NP VP$.

5. There are additional differences in the denotations of the terms *data* and *information* as used by the respective theories and *information* as used by the TLC, but elaboration on them falls outside the scope of this article.
6. Various types of LU can therefore be distinguished, e.g. along the lines of *data categories* in the function theory or *indicator types* in the text theory, but the presentation of such a classification falls outside the scope of this article.
7. For generative (i.e. dictionary-making) purposes the order can simply be reversed.
8. Even in the purest informative referential text, such as a telephone directory, an identity of the sender is encoded, even if it is an identity by default. Consider, however, a telephone directory containing numerous glaring errors with regard to telephone numbers or the alphabetical ordering of surnames and company names. Obviously these are referential and access issues, but the nature and extent of the errors can convey expressive information to the user that can be constructed in the expressive message (representing the image): The publisher is incompetent.
9. The marketing of dictionaries is also a valid research subject in the TLC; in this case the cognate discipline of *marketing communication* is especially informative (as distinguished from the discipline of *interpersonal communication*, which forms the communication theory basis for this article).
10. Many of the symbols (words) in which these characteristics are encoded are in themselves ambiguous, which poses a challenge to the TLC not unlike that faced by, for instance, componential analysis in structural semantics. The same argument applies to the (re)construction of information and messages encoded in and decoded from utterances, since messages are expressed in natural language. A possible solution lies in propositional calculus, but this option will not be explored here.
11. Currently this is a hypothesis that should be tested empirically.
12. FoR_I and FoR_U are indicated by the broken line boxes enclosing *lexicographer* and *target user* respectively in the lexicographical communication model in figure 1.
13. The public pages of the website do not contain any evidence of SPSs. To contribute to the dictionary, a contributor has to log in via his/her Facebook or Gmail account and agree to surrender irrelevant information from these accounts. The author was not prepared to take this step to ascertain whether SPSs exist in the members' area.
14. The statement of the hypotheses does not presuppose that they are or have been proven true.
15. By means of the so-called two-step flow theory from the discipline of *mass communication* (cf. e.g. Steinberg 2007), the TLC can account for the phenomenon that target users are sometimes represented in these types of situation by non-target users or secondary target users. (The exact extent to which the respondents in this case can be seen to represent the primary target user group is not relevant for the purposes of the discussion.)
16. The $p.SQ_U$ is formalised as an illocutionary act to align with the aim.LM_I that offers its answer. $p.SQ_U$ s follow from "raw" potential user questions (= $p.Q_U$ s). For example:

$p.Q_{II-1}$ = What is the part of speech of the lexical item *Buch*?

$p.Q_{II-2}$ = Is the lexical item *Buch* a noun?

$p.Q_{II-1,2} \rightarrow p.SQ_{II-1} \setminus GED$

$p.SQ_{II-1} \setminus GED = ?$ (The word class of lexical item *Buch* is ...)

Between SQ_{II} and $aim.LM_{II}$ in a communication sequence should be inserted *user consultation objective* CO_{II} , thus: $p.Q_{II} \rightarrow p.SQ_{II} \rightarrow CO_{II} \rightarrow aim.LM_{II}$. In the present case $CO_{II} =$ "Find: $aim.LM_{I-2}[LU_9 \setminus LU_7] \in GED$ ". CO_{II} represents the user's intention to navigate the various structures of *GED* to find the needed information (LM) to answer SQ_{II} . For the purposes of this article CO_{II} is irrelevant, as the successful navigation of *GED*'s structures to arrive at the sought LM is assumed. SQ_{II} s are also not discussed in detail.

17. In utilising the TLC for generative purposes, at least one $aim.LM_I$ must answer every $p.SQ_{II}$.
18. Cf. the general typology of lexicographical labels proposed in Beyer (2011).
19. Zero-effects are unlikely to occur in the typical face-to-face interpersonal communication situation, since a receiver's lack of response to a sender's message could in itself represent feedback and therefore an observable effect.
20. Pointed brackets are used to demarcate the limits of complex formal expressions in normal text.
21. This interdisciplinary application allows the TLC to accommodate formal assessment in dictionary pedagogies.
22. It should be noted here that the term (*lexicographic*) *function* has a distinctly different denotation and scope than the identical term in the function theory. What the function theory refers to as *lexicographic function* is accounted for (with adaptations) in the TLC by the term *usage purpose* (cf. Beyer 2013, in preparation). The function theory's concept of *function* is not discussed here; cf. Tarp (2008) for a detailed elaboration. The concept also bears limited resemblance to the concept of *function* in mathematics and set theory, although the same abbreviation is used.
23. The concept of the *usefulness* of a dictionary exists in the TLC (cf. Beyer 2013), but it is not pertinent to the present article.
24. These FoR elements are purely hypothetical and have not been confirmed with either party.
25. This conclusion is based on the fact that LU_{10} is evaluated as *inaccurate*. A specific set of evaluative expressions (such as *accurate, complete, true*) can be assigned to express evaluations of ref.LMs. Using such an expression when evaluating an LU indicates an evaluation of a ref.LM[LU]. Similarly, sets of evaluative expressions can be designated evaluative expressions for exp.LMs and rel.LMs respectively.

Acknowledgement

The author wishes to thank Prof. Ulrich Heid of the University of Hildesheim, Germany, for his constructive criticism on a draft manuscript of this article that has resulted in an improved product. His contribution is part of an on-going scientific collaboration between the University of Hildesheim and the University of Namibia in the framework of the *SeLA* project: *Scientific e-Lexicography for Africa*. Ultimately, however, the author remains solely responsible for the contents.

References

A. Primary literature (dictionary data)

- Macmillan* = *Macmillan Dictionary* online <<http://www.macmillandictionary.com/>> [Accessed 18 August 2014]
Collins English Dictionary online <<http://www.collinsdictionary.com/>> [Accessed 18 August 2014]
Urban Dictionary <<http://www.urbandictionary.com/>> [Accessed 24 July 2014]

B. Secondary literature

- Anderson, L.W. and D.R. Krathwohl (Eds.).** 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. London/New York: Longman.
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Austin, J.L.** 1962. *How to Do Things with Words*. Oxford: Clarendon Press.
- Bergenholtz, H.** 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Beyer, H.L.** In preparation. *Lexicography as Communication*. Unpublished manuscript.
- Beyer, H.L.** 2006. 'n Voorlopige model vir die sistematiese beskrywing van gebruikersvriendelikeheid in woordeboeke. *Lexikos* 16: 46-66.
- Beyer, H.L.** 2009. 'n Teoretiese basis vir kontekstualisering in tweetalige woordeboeke. *Lexikos* 19: 1-22.
- Beyer, H.L.** 2010. Kontekstuele bepalers by die beplanning van skoolwoordeboeke. *Lexikos* 20: 52-72.
- Beyer, H.L.** 2011. 'n Algemene tipologie van leksikografiese etikette. *Journal of Humanities* 51(3): 419-445.
- Beyer, H.L.** 2013. Glosse en vertaalkomplemente in tweetalige woordeboeke. Botha, W., P. Mavoungou and D. Nkomo. (Eds.). 2013. *Festschrift RUFUS H. GOUWS*: 20-38. Stellenbosch: SUN PReSS.
- Bierwisch, M.** 1980. Semantic Structure and Illocutionary Force. Searle, J.R., F. Kiefer and M. Bierwisch. (Eds.). 1980. *Speech Act Theory and Pragmatics*: 1-35. Dordrecht/Boston/London: D. Reidel.
- Bogaards, P.** 2010. Lexicography: Science without Theory? De Schryver, G.-M. (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*: 313-322. Kampala: Menha.
- Brown, G. and G. Yule.** 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Carstens, W.A.M. and K. van de Poel.** 2010. *Teksredaksie*. Stellenbosch: SUN PReSS.
- Cleary, S. (Ed.).** 2003. *The Communication Handbook*. Second edition. Lansdowne: Juta.
- Cruse, A.** 2011. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Third edition. Oxford: Oxford University Press.
- De Schryver, G.-M.** 2012. Trends in Twenty-five Years of Academic Lexicography. *International Journal of Lexicography* 25(4): 464-506.
- Fielding, M. and F. du Plooy-Cilliers.** 2014. *Effective Business Communication in Organisations. Preparing Messages that Communicate*. Fourth edition. Cape Town: Juta.

- Fontenelle, T.** 2011. Lexicography. Simpson, J. (Ed.). 2011. *The Routledge Handbook of Applied Linguistics*: 53-66. London/New York: Routledge.
- Gouws, R.H.** 2012. Who can Really be Called a Lexicographer? *Lexikos* 22: 217-225.
- Grice, H.P.** 1975. Logic and conversation. Cole, P. and J.L. Morgan (Eds.). 1975. *Syntax and Semantics 3: Speech Acts*: 41-58. New York: Academic Press.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta. (Eds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 102-111. Berlin/New York: Walter de Gruyter.
- Janssen, D. (Ed.)**. 1996. *Zakelijke communicatie I*. Groningen: Wolters-Noordhoff.
- Kempson, R.M.** 1977. *Semantic Theory*. Cambridge: Cambridge University Press.
- Kosch, I.** 2013. Expectation Levels in Dictionary Consultation and Compilation. *Lexikos* 23: 201-208.
- Levinson, S.C.** 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Louw, P.A. and H.L. Beyer.** In preparation. The Lexicography of Afrikaans. Hanks, P. and G.-M. de Schryver. (Eds.). In preparation. *International Handbook of Modern Lexis and Lexicography*. Heidelberg: Springer-Verlag.
- Lyons, J.** 1977. *Semantics. Volume 1*. Cambridge: Cambridge University Press.
- Moon, R.** 2014. Meanings, Ideologies, and Learners' Dictionaries. Abel, A., C. Vettori and N. Ralli. (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15-19, 2014*: 85-105. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Piotrowski, T.** 2009. Review of *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography* (Tarp 2008). *International Journal of Lexicography* 22(4): 480-486.
- Ptaszyński, M.O.** 2009. Review of *Bilingual Lexicography from a Communicative Perspective* (Yong and Peng 2007). *International Journal of Lexicography* 22(2): 211-218.
- Rundell, M.** 2012. It Works in Practice but Will It Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusén. (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress, 7-11 August 2012, Oslo*: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Sebeok, T.A.** 2006. Communication. Cobley, P. (Ed.). 2006. *Communication Theories. Critical Concepts in Media and Cultural Studies. Volume 1*: 43-56. London/New York: Routledge.
- Searle, J.R.** 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Steehouder, M. et al.** 1999. *Leren communiceren*. Groningen/Houten: Noordhoff.
- Steinberg, S.** 2007. *An Introduction to Communication Studies*. Cape Town: Juta.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Lexicographica Series Maior 134. Tübingen: Max Niemeyer.
- Van Cuilenburg, J.J., O. Scholten and G.W. Noomen.** 1992. *Communicatiewetenschap*. Second edition. Bussum: Dick Coutinho.
- Van Sterkenburg, P.** 2003. 'The' Dictionary: Definition and History. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography. Terminology and Lexicography Research and Practice* 6: 3-17. Amsterdam/Philadelphia: John Benjamins.

- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim/Vienna/Zürich: Bibliographisches Institut.
- Wiegand, H.E.** 1990. Printed Dictionaries and Their Parts as Texts. An Overview of More Recent Research as an Introduction. *Lexicographica. International Annual of Lexicography* 6: 1-126.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts: An Overview. *South African Journal of Linguistics* 14(4): 134-149.
- Yong, H. and J. Peng.** 2007. *Bilingual Lexicography from a Communicative Perspective*. Terminology and Lexicography Research and Practice 9. Amsterdam/Philadelphia: John Benjamins.

A Functional Analysis of the e-WAT with Specific Focus on the Mobile Version: Towards a Model for Improvement*

André du Plessis, *Department of Afrikaans and Dutch,
Stellenbosch University, Stellenbosch, South Africa*
(ahduplessis@iafrica.com/15612538@sun.ac.za)

Abstract: Lexicography's identity crisis leads to a shift towards electronic lexicographic tools. The rise of mobile phone and smartphone usage in South Africa also creates the opportunity for lexicographers to exploit a new market. This article looks at the current state of Afrikaans mobile dictionaries as an opportunity to delve into this market. The electronic WAT is analysed by primarily applying the function theory to show where user needs must be addressed. This leads to a model for improvement of this and possibly all other mobile dictionaries in South Africa.

Keywords: AFRIKAANS, APPLICATIONS, E-DICTIONARIES, E-LEXICOGRAPHY, ELECTRONIC LEXICOGRAPHY, FUNCTION THEORY, MOBILE DICTIONARY, MOBILE PHONE, LEXICOGRAPHIC TOOLS, SOUTH AFRICAN LEXICOGRAPHY, WAT

Opsomming: 'n Funktionele analise van die e-WAT met spesifieke fokus op die selfoonweergawe: 'n grondslagmodel vir verbetering. Leksikografie se identiteitskrisis lei tot 'n verskuiwing na elektroniese leksikografiese gereedskap. Die groeiende tendens van sel- en slimfone in Suid-Afrika skep ook die geleentheid vir leksikograwe om 'n nuwe mark te ontgin. Hierdie artikel kyk na die onbevredigende toestand van Afrikaanse selfoonwoordeboeke as grondslag vir verdere ondersoek in hierdie sfeer. Die elektroniese WAT word ondersoek deur die funksieteorie toe te pas, om veral aan te dui waar die gebruiker se behoeftes aanspreek moet word. Dit lei tot 'n model vir die verbetering van hierdie en moontlik alle ander selfoonwoordeboeke in Suid-Afrika.

Sleutelwoorde: AFRIKAANS, E-LEKSIKOGRAFIE, ELEKTRONIESE LEKSIKOGRAFIE, E-WOORDEBOEKE, FUNKSIETEOORIE, LEKSIKOGRAFIESE INSTRUMENTE, SELFOON, SELFOONWOORDEBOEK, SUID-AFRIKAANSE LEKSIKOGRAFIE, TOEPASSINGS, WAT

* This article was presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014. It forms part of a master's thesis and is ongoing research.

1. Introduction

Lexicography is at a crossroads. With the advances in technology and the market shift towards electronic tools, the need for digital lexicographic tools is growing (Anderson and Nielsen 2009: 360-362). This means that the areas of theoretical and practical lexicography overlap in the sphere of electronic lexicography. This article will investigate the South African, and particularly the Afrikaans, lack of quality electronic lexicographic tools. The focus will specifically be on mobile dictionaries. A functional analysis of the mobile and online versions of the WAT is undertaken. This analysis creates the opportunity to develop guidelines or recommendations to improve mobile lexicographic tools in South Africa and even abroad.

Extensive international research has been done on e-lexicography and it seems that there is an evident shift towards electronic dictionaries from a theoretical and publishing perspective (cf. De Schryver 2003; Fuertes-Olivera 2009; Heid 2011; Tarp 2012; Lew 2012; Oshima 2013, Smith 2013). South Africa is slowly catching up (cf. Bothma 2011, Prinsloo, et al. 2012, Gouws 2011), but more needs to be done. More projects like SeLA, i.e. the DAAD-funded "Scientific e-Lexicography for Africa", an international collaboration that has yielded some very good research on (South) African lexicography, are needed to kick start the process. This study aims to add to this research output. It must also be added that the goal here is not to criticise good dictionary data, but rather the vessel that the data is presented in.

Please note, for the purpose of this article, the term *mobile dictionary* does not include other electronic devices such as palmtop dictionaries, even though they are mobile electronic tools (See Tono 2009).

2. Research problem and motivation

This study has a dual focus. First, different theoretical principles are studied to determine the theoretical scope to use for an analysis of this nature. This will be briefly explained later on. The physical and technological attributes of mobile phones are taken into consideration as well.

Secondly, the lack of research on Afrikaans mobile dictionaries, as well as the inadequate nature of the current selection of mobile dictionaries available creates the opportunity for further investigation. The nature of one Afrikaans mobile dictionary will be studied. This is important because it is a linguistic resource that can empower many Afrikaans speakers, as well as permeate into other language and culture groups. Thus, the WAT forms the core of this analysis, and is chosen because of its prominence, huge database and availability (as there are no other suitable dictionaries available).

Furthermore, the advances in technology and the fact that there is a global commercial and research shift towards electronic tools, adds motivation to do a study of this nature. Lexicography's current identity crisis, i.e. the split between

printed and digital media, also fuels this fire (Andersen and Nielsen 2012: 360). Tarp (2009: 17) gives three main reasons for this identity crisis:

1. "The theoretical efforts carried out during the last decades and the growing schism between the most advanced theories and the majority of dictionaries produced today;
2. The development of the new computer and information technologies and the present transition from printed to electronic dictionaries;
3. The advent of the so-called information and knowledge society and the growing need for quick and easy access to information."

These reasons essentially act as motivation to delve into electronic lexicographic research. From this perspective it is important that modern and future users be addressed, especially in an era where the mobile phone plays such an immense role in day-to-day activities. This medium cannot be ignored as a possible solution for the need for fast, effective and specialised lexicographic tools.

3. The South African e-lexicography situation

South Africa is a multicultural country with 11 official languages. With the establishment of the National Lexicographic Units (NLU) between 1997 and 2001, a greater emphasis has been placed on language development and dictionary design (Prinsloo 2012: 121). As Prinsloo (2012: 121) indicates, it is an expensive and impractical process to treat every language the same when concerning lexicographic resources. This is reflected in the online dictionaries available to the South African user. Only Afrikaans and English have proper e-dictionaries, whereas Xhosa, Zulu and Sepedi have some form of online dictionaries (Prinsloo 2012: 120-121). It seems that it is not profitable or marketable to design e-dictionaries for the other languages. There are however, current developments in the Afrikaans and Bantu language electronic sphere that must also be taken into account (See Prinsloo 2012).

According to Peyper (2013) there are more mobile phone users than computer users in South Africa. Buys and Malebo (2004: 67-68) indicate that since the inception of cellular communication in South Africa in 1994, mobile phone users increased from 0 to 18 million users in just 10 years. The latest general survey done by Statistics SA (2013) also shows that more than 15 000 households have at least one cellphone owned by the head of the family. That might seem like a small number but it must be noted that other members of the family were not included and that not all households were/can be surveyed. It can thus be argued that if these factors are taken into consideration, the given statistic will certainly increase and would be more akin to that of Buys and Malebo's (2004) research. There is also an already steady foundation of smartphone users in South Africa. In 2011 15% of the population used smartphones and 32% in 2012 (*Our mobile planet* 2013). The latest statistics show that 40% of

the population are now smartphone users (*Our mobile planet* 2013). This is an important statistic, as it shows South Africans' tendency to explore faster and better mobile tools.

When the situation above is scrutinised one logical conclusion is that mobile phone usage in South Africa is likely to grow even faster in the near future. Hence it is imperative that this market be fully exploited and researched from a lexicographic point of view (See for example Winestock and Jeong 2014 for a look at the dictionary app market's growth). Languages that are struggling to adapt to the online sphere or that have a small dictionary market can use mobile dictionaries to start satisfying their users. The NLUs that struggle to find publishers could shift their focus to reaching a new generation of users, whilst instilling a dictionary culture in South Africa and ultimately develop a wide range of users' lexicographic skills in new innovative ways (Prinsloo 2012: 122).

When placing Afrikaans under the microscope, it becomes clear that no fully functional Afrikaans mobile dictionary currently exists. The most Afrikaans dictionaries are mere electronic versions of the printed format. This is not a sustainable model for the future. This also means that these dictionaries were not designed as electronic tools. This has the added implication that the user does not get a usable tool to satisfy his or her lexicographic needs. One such dictionary is unfortunately the e-WAT. However, before this dictionary can be analysed, the necessary theoretical framework must be set out. The following section will give a brief overview of the key aspects used for the analysis.

4. Theoretical analysis of Function theory

The most lexicographic theories are aimed at printed dictionaries, which mean that there is currently a lack of theories for the electronic sphere (Tarp 2009: 23-24). This leads theoretical and practical lexicographers to adapt existing theories for this sphere. One theory that can be adapted is Bergenholtz and Tarp's (2003) function theory as it is fundamentally designed to not be medium specific. This theory focuses on the user and his/her needs. There are four main function types: text reception, text production, translation and cognitive functions (Tarp 2000). The presumption is thus that by using the function theory as part of the theoretical lexicographic approach and applying it on the electronic sphere of mobile dictionaries as part of the practical approach, gaps in the mobile dictionary sphere can be uncovered.

De Schryver (2003: 147) poses an important question that forms the cornerstone of this analysis: "Who accesses what where?" It is essential that the lexicographer take the user of any dictionary into account, as well as where and how the dictionary will be accessed and used. Therefore, a typology of user-needs and usage situations must be drawn up. This leads to the composition of user-profiles, i.e. a specific set of criteria selected or applied to a specific target-user.

The function theory demands that the users' usage situations be typified. There are two main usage situations, namely communicative and knowledge-

orientated (or cognitive) situations. Communicative situations encompass any problems a user experiences with communication, whereas knowledge-orientated situations have to do with situations where the user requires extra information on a specific matter (Tarp 2000: 195).

It is important to note that these functions are always linked to a specific user in a specific context. No two users are alike with regards to their needs, but there are overlapping similarities between users and user-groups which make it possible for dictionary designers to adapt to the majority of needs. Essentially everything in a dictionary is based on functions therefore the user and usage situations form the core of this theory.

5. The Function theory and mobile phones

The electronic medium provides lexicographers with many new and exciting challenges and opportunities. One such opportunity is the fact that a dynamic interaction between user and dictionary can be created. With accessible and manipulable data at the touch of a button, comes new lexicographic challenges all attached to user-needs. As Haas (Bergenholtz and Bothma 2011: 55) states: "the perfect dictionary is one in which you find the thing you are looking for preferably in the first place you look."

This needs to be kept in mind when working with mobile dictionaries. Users want the correct information promptly and easily. They want to solve problems with the least amount of work. Therefore, mobile dictionaries need to be portable and practical problem-solving tools. Bergenholtz and Bothma (2011: 61-62) also link this to "extra lexicographic tool situations":

- Interpreting situations: a user needs help to define or understand signs, e.g. traffic signs, so that he/she can interpret it correctly and follow instructions if need be.
- Operational situations: the user needs help with cognitive or physical operations/processes. The user usually needs guidelines or help with instructions to complete the task.
- Mediating situations: the user needs a portal where he/she can communicate with the lexicographer/developer to log complaints/queries/comments. This is mutually beneficial as it can improve the dictionary on a continuous basis.

Mobile phones, especially smartphones, make these situations tangible. Harnessing the capabilities of the modern smartphone allows dictionary designers to include elements such as quick-links, multimedia, ostensive articles, external sites and effortless communication. This will in turn address the above-mentioned situations. For example, if a user does not understand a certain sign, he/she could be linked to an external site or a multidimensional middle-text with advanced search functions, lists, diagrams, etc. There are still limitations

with cellphones, but lexicographers are not currently exploiting the sophisticated tool that almost every modern user has in his/her possession. It is here that this study aims to make an impact.

6. The WAT

The printed version of the WAT is generally a good dictionary. The e-WAT has potential, but there are some glaring faults that impede effective functional dictionary use.

The WAT is a synchronic and comprehensive multipart dictionary with some diachronic elements. The CD-ROM version was first published in 2003 and the online and mobile versions in 2006, although these versions are in fact mere electronic copies of the printed version. The WAT includes the standard variety and some prominent varieties of Afrikaans. It assumes that its target users are all mother-tongue speakers of Afrikaans, but focuses predominantly on the range between senior high school pupils and academics, in other words more advanced users of Afrikaans (*Elektroniese WAT* 2013). This in itself is a contradiction, as cultural, cognitive and usage-situation factors determine the skill and ability of users, in other words the data presented cannot and does not suffice for all target-users if these said factors are taken into account. Furthermore, a comprehensive dictionary does not have a main function, but the WAT encompasses the functions of text reception, text production and knowledge-orientated functions. The problem with this is the application and manipulation of these functions. A user does not have the choice to select which function they would like to access. This will be explained in detail later on.

Having a good printed version should ideally provide a platform for creating an even better electronic dictionary, as a vast amount of workable data is electronically available, as well as a solid foundation of lexicographic principles that are in place. Unfortunately this ideal does not exist yet, which means that the e-WAT is only a more practical and usable tool than the printed version merely due to users having all the parts in one environment with quick access to lemmas through a search function.

7. Technical Analysis of the e-WAT

Before the mobile version can be dissected, a quick look at the online version is needed. These two versions are very similar and the e-WAT stems from the online one, so analysing the online one allows for a more comprehensive analysis all-round. This analysis, as mentioned, is done from a functional perspective, i.e. taking the user's experience with the dictionary into consideration. This also means that effectiveness and usability is ultimately tested.

Figure 1 shows the online version's homepage and basic search function. There are also links to the mobile version, advanced searches, outer-texts and information about the WAT. Figure 1.1 shows the online WAT's advanced

search option. This is an acceptable search mechanism but still needs refinement, especially concerning fuzzy logic and user guidance or error. There is also no way to sign out of the dictionary or to start a new fresh search. Moreover, there is no way to go back to previous searches; unless the web browser's back button is used which then requires the user to refresh the web browser. It would be sensible to have a link that allows users to see their previous searches or have a back button available in the dictionary itself. These might be small things, but they cause frustration and needs to be streamlined to enhance the dictionary's effectiveness.

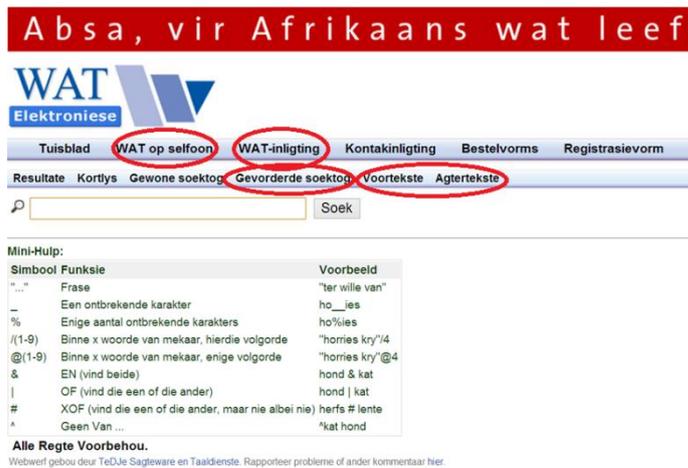


Figure 1: The online WAT

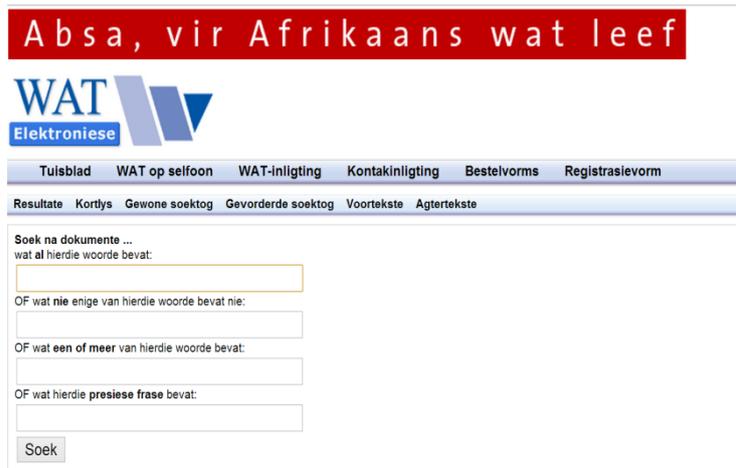


Figure 1.1: Advanced search options

The problem is that the hit-list and shortlist don't really offer the user much lexicographic guidance and serves no real functional purpose seeing that a user still has to go into that article to find what he/she is looking for. Consequently, it is unclear why the user is presented with these lists. It would be more sensible to offer manipulable data so that the user can individualise his/her WAT experience. As figure 1.3 shows, the only data that is otherwise clickable in the article is when a cross-reference is given. Online dictionaries should be interactive environments where the user can select the exact data he/she is looking for. Not being able to select the functions, data-types or usage-situations, negatively affect the user's ability to use the dictionary effectively. A user must also be allowed to create his/her own changeable profile that allows for different search options depending on the type of user. For example, an advanced user (semi-expert) in law may not need text production information about a certain legal term or phrase, but just wants to check a single meaning to make sure he/she is correct in using the word/phrase.

Upon further inspection of the outer-texts, it seems that user guidance as a whole needs to be addressed in this version. The user is confronted with outer-texts that are not linked to the functions of the dictionary they are using. The online WAT's outer-texts give users the same information and usage-guide found with the CD-ROM version (a version that runs on Folio Views and has different search routes). Figure 1.4.1 and 1.4.2 illustrates this incongruity. This is of no help to any user that uses the online or mobile version as there are functions, processes and search routes that are not applicable to the online version. Furthermore, a lot of irrelevant information is presented. It is understandable that the dictionary contributors should get credit, but the layout of the guidance text places this ahead of crucial usage information. An online user wants a simple and quick guide to understanding and using the dictionary. The online WAT fails in this regard. It is imperative that further research be done on the online WAT, as this can permeate into the mobile version and into other South African online dictionaries. The above-mentioned critiques are some of the areas that, once improved, can be applied to the mobile version to enhance its functionality too.

'n Eenvoudige soektog deur middel van die *Simple Search Dialog Box*

Die *Simple Search Dialog Box* is aan die bokant van die skerm.

Stap 1: Tik die woord (bv. horries) of 'n frase waarvoor jy wil soek (tussen aanhalingstekens, bv. "horries kry") in die *Simple Search Dialog Box*.



Stap 2: Klik *ENTER* om die soektog te aktiveer. Die program sal jou neem na die eerste voorbeeld van jou soektog.

Stap 3: Gebruik die *Vorige resultaat* (*Previous Hit*)- en *Volgende resultaat* (*Next Hit*)-knoppies op die *Simple Search Dialog Box* om deur die soekresultate te beweeg.

Snelskakel: F4 neem jou na die volgende resultaat en Shift+F4 na die vorige resultaat.

Figure 1.4.1: The CD-ROM versions user guidance given in the online version

Die Advanced Query Dialog Box

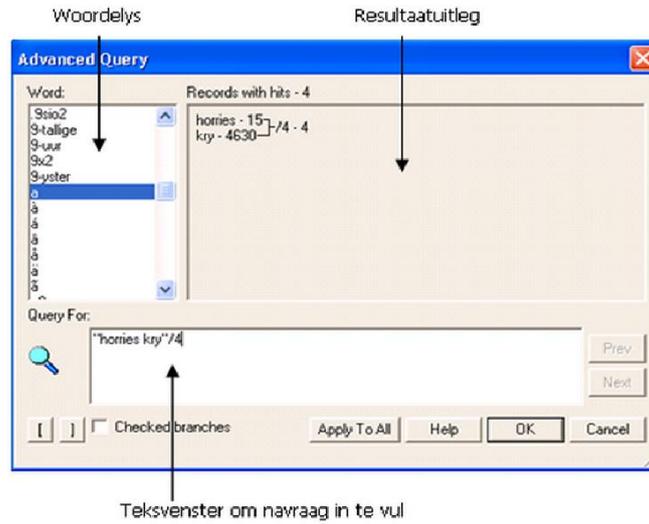


Figure 1.4.2: The advanced query dialog box function. Only available in the CD-ROM version but presented in the outer-text of the online version

7.1 The mobile WAT

The mobile version also suffers from a lack of functionality and effective use of available technological attributes. As with the online version, this version is also a digital copy of its printed counterpart. Fundamentally, it is almost a carbon copy of the online WAT. This means it has many of the same problems as well as a few added obstacles. Figure 2 shows the mobile home screen:

[WAT-Selfoon Tuisblad](#)

Resultate

Alle Regte Voorbehou

Figure 2: The mobile WAT

It is clear from Figure 2 that this version offers even less help to its users than

the online one. This very basic homepage is all that a user is confronted with. Furthermore, this version's search function is very limited. Only single word items can be searched for and no advanced searches can be undertaken whatsoever (Botha 2013). No outer-texts, no links and no help or any other options are given. Even if the user is an advanced user, no dictionary should assume that its users do not need usage guidance. It seems that this version leaves the user to fend for him/herself when navigating or using the dictionary. At first glance, it seems this version is thus not very user-friendly.

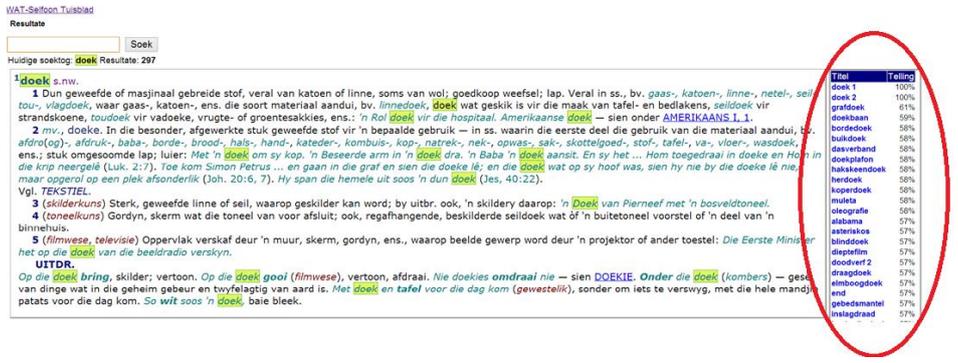


Figure 2.1: A basic search done on the mobile WAT

When a basic search is done, as figure 2.1 shows, the dictionary delivers the results in the same manner as the online version. This is problematic. Again, it must be stated that using the same model as the printed version for an electronic environment, is not sustainable or practicable. For example, figure 2.1 shows the ineffective hit-list, the same layout as the online version and a very static digital environment. The developers also failed to realise that a cellphone has certain limitations with regards to screen layout, but have processing power that allows for innovative ways to display and manipulate data. This means that the layout and structure of the dictionary need to be addressed and the implementation of technological possibilities should be considered. Figure 2.2 echoes this point:

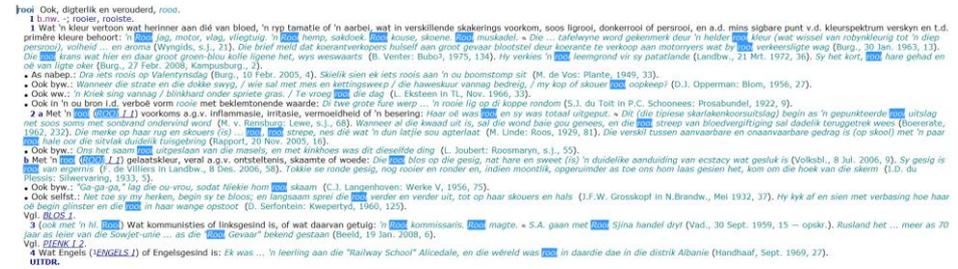


Figure 2.2: Problematic search results

The search done in figure 2.2 shows the results for the article "rooi" (red). Note that this is not even the full article. It is understandable that due to the WAT's encyclopaedic nature, comprehensive articles are common practice. The problem is not the data but rather the way it is packaged. Therefore an article like "rooi" leads to the following conclusion about the mobile WAT's structure and layout: the user is left to his/her own devices to decipher and sift through a huge mass of information, which in turn frustrate and confuse the user and impairs functionality and usability of the dictionary. Trying to navigate through this dictionary on a cellphone screen is no easy task and by simplifying and adapting the dictionary for mobile use, will greatly improve its functionality. As previously highlighted, the modern dictionary user wants a tool that is simple, effective and adaptable. This means that the WAT must adhere to these needs on a structural and functional level. Having a static layout with an overload of data makes it increasingly difficult to use the dictionary.

As with the online version, not being able to manipulate the data or create a custom profile inhibits the user from satisfying his/her individual needs quickly and easily. The data must be presented in such a way that the lexicographer and user have control over the display and accessibility of data. The mobile WAT does not make use of the technological possibilities on offer to improve these shortcomings. Having the option to create user profiles could be a way to enhance the user's experience, as well as the way data is presented. Thus, innovative access and article structures must be implemented. For example, a manageable home screen with different options and settings that allow users to create an individualised version of the dictionary would vastly improve its structural aspects. This can also be achieved by having tick-boxes for the discrimination of functions or data-types; incorporating a logically set out interface and removing the hit-list and index.

As with the layout, the search functions also need rethinking. Modern smartphones can provide quality products that have lightning quick smart search engines, for example a search function that stores searches, autocompletes searches or aids users by providing suggestions, similar to Google's *Did you mean* function. As the WAT has such a wide range of users with different language skills, one aspect that also needs addressing is the occurrence of user error in searches. Many users use a dictionary for text production and have trouble spelling. By implementing a spellchecker in the search engine will improve user control and user guidance.

Most mobile phones also have the capacity for multimedia elements. The printed version has some ostensive elements, although they are not used consistently. The mobile and online sphere offers the chance for developers to incorporate different multimedia elements consistently, by creating an interactive dictionary. This in turn would also enhance the user's experience, allow for the knowledge-orientated function of the dictionary to be modernised and improve on the functionality of the dictionary. The use of different multimedia links should thus be taken into consideration. This can be done in the form of

video clips, ostensive defining and multi-dimensional cross-referencing, i.e. having additional clickable links that either displays a diagram/photo/picture/video or that takes the user to an external site where more information is given. Pop-ups can also be used where either with additional, grammatical or syntactical information is given should the user require it. This would also aid users in extra lexicographic tool situations.

As Bergenholtz and Bothma (2011: 61-62) states: a dictionary is an information tool to be used by a specific set of users. The WAT is a good information tool but lacks accessibility and usability. The e-versions compensate by being more accessible, but not functional. A new brand of techno-savvy user is on the rise and if the WAT wants to make an impact on these users, a mind-shift needs to be made. According to Gouws (2013) the current crop of printed WAT users will eventually die out and be replaced by the modern user with a know-how perspective of all things mobile and electronic. Even Botha (2014), the editor of the WAT, states that there is an increase in the sales of electronic versions and that mostly older people buy the printed version. This means that eventually those older users will disappear, which leaves the new generation of users, a generation with different expectations and needs, with a product that cannot satisfy their needs or meet those expectations.

Thus for the e-WAT to be sustainable and marketable as a practical and proper lexicographic tool in the long-term, a division of resources must be allocated. In other words the current versions and future parts should be reworked to be more online or cellphone compatible. Funding is a serious problem and the WAT has struggled with service providers and licensing costs for the past two years (Botha 2014), which has slowed production. This is a problem that can potentially be solved by setting the e-versions as a priority, as it could generate a new market of modern users. This also allows for the updating of older parts and having a dictionary that is continuously updated, which in turn continuously satisfy user needs. The mobile and online version should also be promoted to flagship projects and should not only be seen as part of the printed version's progress.

8. Recommendations

This section will provide a few recommendations that stem from the WAT's analysis, which can be applied to mobile dictionaries in general. Lew's (2012: 344) statement is valuable to keep in mind when developing recommendations: "dictionaries should be able to satisfy [user] needs in a short time and with a degree of detail. [T]he data has to be presented in a form that is maximally comprehensible."

Firstly, it must be emphasised that South Africa does not have the infrastructure to support dictionary development like Great Britain, Germany or Japan. We also have a vast array of multicultural users. As Prinsloo (2012) proposes for online dictionaries in South Africa, our future mobile dictionaries

should also initially be monofunctional. Only when we perfect monofunctional dictionaries and the standard for mobile dictionaries has been set, can we start designing polyfunctional dictionaries. This proposal is especially aimed at new dictionaries or at languages that do not have proper standardised dictionaries.

The effectiveness of any product is usually analysed in conjunction with the products relationship with its users (Heid 2011: 292). This goes for dictionaries too. The differences in user needs will ultimately play a part in the fluctuation of effectiveness. Therefore Heid (2011: 294) proposes some criteria whereby effectiveness is measured. Even though these criteria are for most machine interactions, they have been specifically adapted to suit electronic dictionaries:

- Stability for the task: the dictionary interface should be able to facilitate a range of tasks as set out by the user. An interface should streamline communication between the dictionary and the user.
- Self-descriptiveness: the dictionary interface should be clear and informative, give clear feedback and data should be visible and recognisable.
- Controllability: user interaction is crucial. Users should have control over their dictionary interface.
- Conformity with user expectations: the dictionary should behave in a way the user expects, i.e. should behave as a dictionary should and not be unfamiliar to users.
- Error tolerance: user error must be expected and processes must be in place to guide users when this occurs. Reversals should be done indeed necessary.
- Suitability for individualisation: customisation or individualisation should be allowed. A user must make the dictionary his/her own.
- Suitability for learning: the dictionary should be quick and easy to master, with adequate guidelines and mediating functions in place.

It is very important that a dictionary developer takes these criteria into consideration when designing a mobile or online product, as this has a significant effect on the usability and marketability of the dictionary.

Heid's criteria fits well within the framework that Fuertes-Olivera and Niño-Amo (2011: 172-173) drew up for online dictionary design. It is also possible to adapt some of these guidelines for mobile dictionaries. As with the Heid's criteria, my supplementary ideas have been amalgamated with Fuertes-Olivera and Niño-Amo's guidelines:

- Smart Searches: A mobile dictionary should be able to do smart searches, similar to Google's search engine. Different search routes or functions should be available. Different articles/lemmas/data-types should be searchable in different ways. For example being able to highlight words in a defi-

inition and being able to search for said text.

- Interactivity: Users must feel part of the dictionary. The dictionary should be a familiar environment that highlights user interaction by means of various personalisation options, clickable links, user input and/or guidance. They should also not be alienated by confusing data-types/styles/presentations. They should also have the freedom to give their input to enhance the dictionary.
- Layout: User-friendliness is essential. The dictionary must be pleasant to look at, incorporate different technological measures to ensure that it is fast, simple, aesthetically pleasing and logically set out. Everything should be clearly labelled and must feel inviting to use. Bothma's (2011: 83) ideas for online dictionaries are also valid here: mobile dictionaries must be designed in such a way that users can navigate effortlessly through the macro- or microstructure. They must not be bombarded with options, details or links that crowd the screen or that cannot close/be minimized. There should be a main menu that is easy to understand and follow, and which can be returned to at any time.
- Multimedia possibilities: Multimedia should be incorporated, especially as a variation on data-types or as supplement to existing articles. The use of ostensive defining, video clips or even sound clips can be utilised to present users with extra information on articles if needed. Sound clips can be very handy for pronunciation purposes, as users can at least hear the standard variety of pronunciation of a word.
- Familiarity: The virtual environment should not be too different from a printed dictionary, i.e. the dictionary should still be recognisable as a dictionary. Articles can be given in innovative ways but should still meet the average user's expectations. Adverts and other distracting elements should be avoided, as this limits the screen space of the dictionary and presents an overload of data.
- Extra software functions: Functions that are linked to the mobile phone, such as copy-and-paste, highlighting text, minimizing or signing in/out, should be taken into consideration as they offer extra functionality and usability to the dictionary.
- Updating: Updates and corrections are needed, especially if users have asked for them. Updates can be expensive or time-consuming, but they are a necessary part of improving dictionaries. Updates should be done regularly (once a week at least), but users should be able to choose whether they want the update or not. Apps that continuously update are very irritating and consume precious network data, which is problematic for users that only have a set amount of data each month. The dictionary should still be functional even if it is not updated or it should be downloadable with an offline-mode.

- Search results: Firstly, search results must be readable and fit to the size of the phone's screen. Only relevant information should be presented. The information/data should be logically set out. As mentioned, advanced searches should be a given, but they should be manipulable and be usable in search results too, for example users should be able to dictate what criteria they want their results to be sorted by (theme, word length, collocations, etc.).

The analysis of the WAT emphasises the lack of customisation that e-dictionaries seem to have. One area that can improve this shortcoming is to introduce user profiles. As Bothma (2011: 84) states: "This can help to reduce information overload and provide the user with customized information tailored to his/her specific needs in a specific situation." This implies that users will find relevant data quicker if they have a profile set to their specifications. Thus, another recommendation can be added to the above-mentioned:

- User profiling: Users should be able to sign into a dictionary and create a custom profile that can be adapted to their needs. This in turn will emphasise the adaptability of mobile dictionaries as practical tools. The user profile must indicate: the user type (expert, semi-expert or layman — this can refer to dictionary usage knowledge or language proficiency) and their function or data-type preference.

Another area that needs to be addressed is that of guidance in text production, especially in translation situations. Prinsloo, Heid, Bothma en Faaß (2012: 292) have already looked at this in online dictionaries, but it pertains to mobile dictionaries too. This can be very useful for creating monofunctional translation dictionaries, for second or foreign language speakers. Giving user guidance in the form of pop-ups would work well. These pop-ups could contain grammatical, syntactical or morphological information about the source language (Prinsloo et al. 2012: 296). This is not the only method, as innovative markers could also be used. These markers could be in different colours and shapes, and would be used to warn users about the use of loaded or confusing words, homonyms, homophones, etc. For example an interactive red flag that when clicked gives a concise explanation and an example to illustrate the usage situation.

These guidelines or recommendations are just a few key aspects that need to be addressed if mobile dictionaries are to improve. One of the biggest problems that currently exist in South African lexicography is the use of printed dictionary structures for online and mobile user. Dziemianko (2012: 321) makes a good point when she says that online and mobile dictionaries do not have to have rigid structures like printed dictionaries. The fact that "layered presentations" can be used and that dynamic data-presentation is possible, as well as the fact that there are no real limits to storage space, should give these mediums authority in the modern era (Dziemianko 2012: 321). The printed diction-

ary sphere should however not be ignored as it too can benefit from the improvement of online or mobile dictionaries.

9. Conclusion

This article looked at the e-WAT as a guinea pig for the application of some lexicographic theories as a starting-point to improve Afrikaans and other cellphone dictionaries. Many of these recommendations are not applicable in the short-term, but need to be addressed if the WAT wants to survive in a very fast changing lexicographic landscape. The current e-versions are not as effective as they can be. There are still too many shortcomings for it to be a sustainable and practically usable electronic lexicographic tool. The WAT is a very good dictionary and they do an outstanding job in setting the standard for lexicographic practice in South Africa, but the fact remains that the shift towards electronic mediums will greatly affect them too in the not-too-distant future.

Creating a tool that encompasses the technological possibilities available and that satisfies the new generation of users' needs, will bode well for South African lexicography and might even attract a completely new group of users to the WAT. This is the perfect opportunity to fill the void that currently exists in the Afrikaans cellphone market.

A range of recommendations is made that can be incorporated into mobile dictionaries. It is obvious that not all of these recommendations are applicable on every dictionary or on one single dictionary, but a combination of these guidelines should be taken to heart when designing mobile lexicographic tools for the future.

In conclusion, a lot of research is still needed on mobile and online dictionaries, and as with any scientific process, a period of trial and error lies ahead. This study shows that the mobile version of the WAT is not yet a truly functional electronic tool, but by applying some of the recommendations made for improvement, it will become the basis for a whole range of other mobile dictionaries in South Africa or Africa. There are, however, also other problems that South African lexicography is faced with. Many socio-economic factors impede the implementation of sustainable reference sources in South Africa. Our multicultural population can also cause some obstacles for the development of online or mobile dictionaries. More research needs to be done in this regard.

10. References

- Anderson, B. and S. Nielsen. 2009. Ten Key Issues in Lexicography for the Future. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009: 355-363.
- Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopaedias Today, Lexicographical Tools Tomorrow*. Linguistic Insights. Studies in Language and Communication 90. Bern: Peter Lang.

- Bergenholtz, H. and T.J.D. Bothma.** 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- Botha, W.F.** (wfb@sun.ac.za), 15 August 2013. Re: Selfoonwoordeboek. E-mail to A. du Plessis (15612538@sun.ac.za).
- Botha, W.F.** 2014. Personal Interview, 27 March 2014, Stellenbosch.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in an Online Environment in Response to User Needs. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 71-102.
- Buys, A.J. and L.M.M. Malebo.** 2004. Managing Rapid Diffusion: The Case of Cellular Communications in South Africa. *SA Journal of Industrial Engineering* 15(2): 67-78.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Dziemianko, A.** 2012. On the Use(fulness) of Paper and Electronic Dictionaries. Granger, S. and M. Paquot (Eds.). 2012: 319-337.
- Elektroniese WAT.* 2013. [Online] <http://www.woordeboek.co.za>.
- Fuertes-Olivera, P.A.** 2009. The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-language Internet Dictionary. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009: 99-131.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.).** 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum International.
- Fuertes-Olivera, P.A. and M. Niño-Amo.** 2011. Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés Español de Contabilidad*. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 168-186.
- Google.** 2013. *Our Mobile Planet*. [Online] <http://think.withgoogle.com/mobileplanet/en/>.
- Gouws, R.H.** 2011. Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 168-186.
- Gouws, R.H.** 2013. Die geskiedenis van die WAT. Class Lecture (Afrikaanse Leksikografieteorie 771). Stellenbosch: Stellenbosch University.
- Granger, S. and M. Paquot (Eds.).** 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Heid, U.** 2011. Electronic Dictionaries as Tools: Toward an Assessment of Usability. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 287-304.
- Lew, R.** 2012. How Can We Make Electronic dictionaries More Effective? Granger, S. and M. Paquot (Eds.). 2012: 343-361.
- Oshima, J.** 2013. The Digital Dictionary: Classroom Use in Japan. Paper read at the *8th International Conference of the Asian Association for Lexicography (ASIALEX 2013)*, 20-22 August 2013, Bali, Indonesia.
- Peyper, L.** 2013. Mobile Phone Usage in SA.. *Finweek* [Online] <http://finweek.com/2013/01/22/mobile-phone-usage-in-sa/>. [Accessed: 24 September 2013].
- Prinsloo, D.J.** 2012. Electronic Lexicography for Lesser-resourced Languages: The South African Context. Granger, S. and M. Paquot (Eds.): 119-144.
- Prinsloo, D.J., U. Heid, T. Bothma and G. Faaß.** 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.
- Selfoon-WAT.* 2013. [Online] <http://sel.woordeboek.co.za>.
- Smith, W.** 2013. Pharos Dictionaries. Personal Interview, 26 August, Cape Town.
- Statistics South Africa.** 2014. *General Household Survey 2013*. Pretoria: Statistics South Africa.

- Tarp, S.** 2000. Theoretical Challenges to Practical Specialised Lexicography. *Lexikos* 10: 189-208.
- Tarp, S.** 2009. Beyond Lexicography: New Visions and Challenges in the Information Age. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009: 17-31.
- Tarp, S.** 2012. Theoretical Challenges in the Transition from Lexicographical p-Works to e-Tools. Granger, S. and M. Paquot (Eds.). 2012: 107-118
- Tono, Y.** 2009. Pocket Electronic Dictionaries in Japan: User Perspectives. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009: 33-67.
- Winestock, C. and Y. Jeong.** 2014. An Analysis of the Smartphone Dictionary App Market. Tono, Y. (Ed.). 2014. *Lexicography Journal of ASIALEX* 1(1): 109-119.

A General Lexicographic Model for a Typological Variety of Dictionaries in African Languages

Gertrud Faaß, *Department of Information Science and Natural Language Processing, University of Hildesheim, Germany (gertrud.faaß@uni-hildesheim.de)* and *Department of African Languages, University of South Africa, Pretoria, South Africa*

Sonja E. Bosch, *Department of African Languages, University of South Africa, Pretoria, South Africa (boschse@unisa.ac.za)*

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

Abstract: So far, there have been few descriptions on creating structures capable of storing lexicographic data, ISO 24613:2008 being one of the latest. Another one is by Spohr (2012), who designs a multifunctional lexical resource which is able to store data of different types of dictionaries in a user-oriented way. Technically, his design is based on the principle of a hierarchical XML/OWL (eXtensible Markup Language/Web Ontology Language) representation model. This article follows another route in describing a model based on entities and relations between them; MySQL (usually referred to as: Structured Query Language) describes a database system of tables containing data and definitions of relations between them. The model was developed in the context of the project "Scientific eLexicography for Africa" and the lexicographic database to be built thereof will be implemented with MySQL. The principles of the ISO model and of Spohr's model are adhered to with one major difference in the implementation strategy: we do not place the lemma in the centre of attention, but the sense description — all other elements, including the lemma, depend on the sense description. This article also describes the contained lexicographic data sets and how they have been collected from different sources. As our aim is to compile several prototypical internet dictionaries (a monolingual Northern Sotho dictionary, a bilingual learners' Xhosa–English dictionary and a bilingual Zulu–English dictionary), we describe the necessary microstructural elements for each of them and which principles we adhere to when designing different ways of accessing them. We plan to make the model and the (empty) database with all graphical user interfaces that have been developed, freely available by mid-2015.

Keywords: AFRICAN LANGUAGES DICTIONARIES, LEXICOGRAPHIC MODEL, MACROSTRUCTURE, MICROSTRUCTURE, ACCESS STRUCTURE, ISO24613:2008, MYSQL, MULTIFUNCTIONAL LEXICAL RESOURCE

Opsomming: 'n Algemene leksikografiese model vir 'n tipologiese verskeidenheid woordeboeke in Afrikatale. Tot dusver bestaan daar min beskrywings oor hoe

om strukture te skep wat daartoe in staat is om leksikografiese data te berg. ISO 24613 2008 is een van die mees onlangse sodanige beskrywings. Nog een, naamlik dié van Spohr (2012) wat fokus op die ontwerp van 'n gebruikersgerigte multifunksionele leksikale bron, gebruik die voorstellingsmodel XML/OWL (eXtensible Markup Language/Web Ontology Language) wat in beginsel hiërargies is. In hierdie artikel word 'n ander roete gevolg om 'n model te beskryf wat gebaseer is op entiteite en hul onderlinge verhoudinge. MySQL (gewoonlik na verwys as Structured Query Language) beskryf 'n databasisstelsel van tabelle wat data bevat en definisies van hulle onderlinge verhoudinge. Die model is ontwikkel binne die konteks van die projek "Scientific e-Lexicography for Africa" en die databasis wat saamgestel word, sal met behulp van MySQL toegepas word. Die beginsels van die ISO-model asook dié van Spohr word gehandhaaf maar wel met een groot uitsondering in die implementeringstrategie: die lemma is naamlik nie sentraal in die bewerking nie, maar wel die beskrywing van betekenisonderskeidings — alle ander elemente, met insluiting van die lemma, is afhanklik van die beskrywing van die betekenisonderskeidings. Hierdie artikel bespreek ook die leksikografiese datastelle wat aangebied word en hoe hulle uit verskillende bronne versamel is. Aangesien dit ons doel is om verskillende prototipiese internetwoordeboeke saam te stel ('n eentalige Noord-Sotho woordeboek, 'n tweetalige aanleerderswoordeboek Xhosa-Engels en 'n tweetalige woordeboek Zoeloe-Engels) bespreek ons die noodsaaklike mikrostrukturele elemente vir elkeen van hierdie woordeboeke en watter beginsels gevolg word om verskillende maniere te ontwikkel om toegang tot hierdie woordeboeke te verkry. Die plan is om die model en die (leë) databasis met al die grafiese koppelvlakke wat ontwikkel is teen die middel van 2015 gratis beskikbaar te stel

Sleutelwoorde: AFRIKATAALWOORDEBOEKE, ISO24613:2008, LEKSIKOGRAFIESE MODEL, MAKROSTRUKTUUR, MIKROSTRUKTUUR, MULTIFUNKSIONELE LEKSIKALE BRON, MYSQL, TOEGANGSTRUKTUUR

1. Introduction

This article is concerned with the design of a lexicographic model, that is, a model of a data structure capable of storing lexicographic data, which will subsequently be used to compile several types of prototypical dictionaries for a selection of African languages¹. We keep in mind that there are no hard and fast rules for any typological model, but rather that different types of dictionaries may have certain features in common (Gouws and Prinsloo 2005: 45). In the last few years, several such lexicographic data collection models were published; the most general of all is the ISO standard for lexicography (ISO 24613:2008). This "Lexical Markup Framework²" (LMF) builds the background for several existing lexicographic data collections. A data collection model is not a database as such, but is defined as a standoff-XML-formatted framework of a number of files plus several external sources, each describing a different aspect of the dictionary that is compiled from them. For example, general data such as language or language coding is included, but also microstructural data related to lemma signs, such as information about its part of speech or its orthography. Concerning the possibilities to connect with other sources of

information, we agree with Spohr (2012: 23) who states that although LMF describes itself as interoperable, "it remains rather vague on its application in the various contexts, and in particular of its application in human usage situations".

Spohr's general graph-based formalism (Spohr 2012) can indeed be seen as an implementation of the LMF data model. His lexical resource, implemented in a graph based OWL model, is based on a typed formalism, similar to the adaptations the WWW is taking up to become the new Semantic Web (Spohr 2012: 38). Spohr places the lexeme in the focal point of the database, linking it for instance to its forms and senses (*ibid.* p. 68). He nevertheless states that "ideally, we would like senses to be the primary lexical entities, as all kinds of lexical relations seem to be defined between senses" (*ibid.* 67). Spohr, however, then argues against this concept saying that beginning with the item giving the sense (i.e. the item giving the paraphrase of meaning), it would not be possible to fill all other dependent fields, especially when acquiring lexicographic data from corpora (*ibid.* 68). This issue will receive further attention in section 6.3.

We want to mention two further publications here, which describe a lexicographic database or data collection model for generating online dictionaries in particular. A database that supplies several dictionaries for specific purposes with data is described by Bergenholtz and Bergenholtz (2013). In their article titled "One database, four monofunctional dictionaries", the kind of model that was utilized is unfortunately not mentioned. However, they do point out some items defined for the resulting database, as well as the fact that the compilation of several online dictionaries from one database, calls for a number of issues concerning its access features to be taken into account — see also our section 5 below.

Bosch, Pretorius and Jones (2007) propose a model for machine-readable lexicons, not only for the South African Bantu languages, but for the Bantu language family as a whole. The data model in the form of an XML DTD is intended to include all linguistic information of the languages in question and "provides flexibility and handles the various representations specifically applicable to Bantu languages, thereby making it applicable to diverse uses of machine-readable lexicons" as language resources for use in large-scale HLT/NLP applications. Only a fragment of the DTD is presented in the publication.

The majority of articles concerned with online dictionaries, however, refers to their visual representation (e.g. Prinsloo 2010 which is related to their implemented access strategies), others are concerned with the acquisition of data to populate lexicographic databases (e.g. L'Homme 2012 and Scholze-Stubenrecht 2013).

The research for this paper resides within a project entitled "Scientific e-Lexicography for Africa (SeLA)³" (i.a. described by Heid 2012), and it is carried out by the University of Hildesheim (Germany), the University of South Africa and the University of Pretoria, Stellenbosch University (South Africa), and the

University of Namibia in Windhoek (Namibia). The project intends to combine all of the above-mentioned issues: (1) designing a prototypical multifunctional database with the aim of compiling several monofunctional electronic dictionaries for the African languages; (2) solving the problem of data acquisition for resource-scarce languages; (3) defining "exactly which types of lexicographic data from the fact collection need to be selected in order to satisfy a given user need, as well as in deciding in which way such data have to be ordered and formatted (presented) for users with a given background and a given type of need" (Heid 2012: 438).

In the SeLA project, we are concerned with a multilingual African language data collection to be used for lexicographic purposes which we will store in a MySQL database. For the time being, the aim is not to compile comprehensive dictionaries from the database. Seeing the final implementation as a prototype, we plan to use this database for several other purposes, for instance, as part of intelligent Computer Assisted Language Learning (iCALL) software.

We consider it necessary to strictly differentiate between the database, which should be flexible, in other words, open to internal and external resources (so far unknown) to be added in the future, and the presentation of the (internal and external) data to the users, which depends on their requirements (see section 5). We also foresee access to a prototypical Natural Language Processing (NLP) machine performing morpho-syntactic analyses.

The database model is to be implemented with a MySQL database. Such a database may consist of (1) content tables containing the data itself, (2) relational tables linking data items with one another, and (3) tables generated from the data and their relations which are used for a faster access. One might wonder why we do not use XML/OWL, like the most up-to-date data collection models described above. Besides the fact that the SeLA team lacks the capacity to develop a full-scale Dictionary Writing System (DWS) or to make use of one to compile a full-scale dictionary, we consider a populated MySQL database implementation as equal to a standoff XML system. In both systems, all necessary data items can be described and a number of types of relations between those data items can be modelled. SQL, however, additionally allows for a fast and easy implementation without the need for DTDs, XML-editors or (commercial) Dictionary Writing Systems. Moreover, together with phpMyAdmin⁴, an online dictionary and the necessary maintainer facilities are speedily and simply implemented with a few PHP scripts. Another point of consideration is that most of the data will be imported from existing resources, which will populate the fields of the database only partially. The task of filling the gaps and generating full-scale dictionaries must be postponed to a later stage. To use MySQL for a start, does not imply that XML/OWL will not be used in the future. In such a case, the means will be found to fill the database with sufficient data to compile comprehensive dictionaries, and porting one system to the other will indeed be possible.

In summary, we describe a lexicographic model in this article which should

fulfil various requirements: (1) it should be open to a number of lexicographical functions as several different monofunctional online dictionaries will be compiled from it; (2) it should cover the specific linguistic phenomena of the languages belonging to the Bantu language family; and (3) concerning data acquisition — as we will need to populate the database with any relevant data that can be collected semi-automatically — the database should be tolerant of missing data items, even if they are considered essential for producing a dictionary. Furthermore, we will describe our current approach towards data acquisition and data accessibility.

2. Aims

Our aim as part of the SeLA project is to design and develop a lexicographic database that will contain multilingual data of three of the official African Languages of South Africa (i.e. Zulu, Northern Sotho and Xhosa). For some of these data sets, translation equivalents of South African English will be stored too. The data of other African languages, as well as Afrikaans, are foreseen to be added at a later stage. We begin by developing a database model, with the aim of fulfilling all the requirements to describe the language items thoroughly, while taking into account the languages in question and the external resources that are currently available. We take Spohr's (2012) data collection model into account too; however, as Spohr has suggested, we focus our attention on the polysemous senses of a word — the above-mentioned disadvantages (see section 95) only play a minor role for us, as is the case with the languages concerned, there are only few resources available which would allow for an automated filling of the database — most data will have to be added manually. The database will be utilised to compile a typologically diverse collection of prototypical monofunctional dictionaries (however, with few data sets), of which the majority are planned to be bilingual. Hence, we look at requirements of a good outer and inner access structure (see section 5), resulting in the design of different dynamic graphical user interfaces (GUIs) to be developed.

We will then examine ways and methods to import available external resources (the respective plans are described in section 6). Lastly, we plan to bind the resulting database into a language portal, a framework of lexicographic and other resources. We foresee linking it with other dictionaries, corpora, or other databases containing linguistic data, such as the ontology database of the part-of-speech items of Zulu and Northern Sotho described by Faaß, Bosch and Taljard (2012) or the e-learning tool "eZulu dictionary of possessives" assisting learners of the language in acquiring knowledge about producing possessives structures in Zulu, described by Bosch and Faaß (2014).

Setting the aims as described above, we need to examine aspects regarding macrostructures and microstructures of the foreseen dictionaries. On this basis, the data model can then be designed.

3. Aspects regarding the macrostructure and microstructure

3.1 Macrostructural elements for Bantu language dictionaries: a challenge of lemmatisation

The agglutinating nature of the Bantu languages that goes hand in hand with a complicated nominal and verbal derivation system, indeed poses challenges for lemmatisation (Gouws and Prinsloo 2005: 67). Different approaches to lemmatisation, the main one being word versus stem lemmatisation in the case of nouns and verbs, play an important role in dictionary compilation.

Because of the conjunctive writing system of Zulu, whereby parts of speech are written together, even full sentences may appear as one orthographic word. The sentence *bazokubona* "they will see it", for example, consists of several morphemes; *ba-* (subject concord of noun class 2) *-zo-* (future tense marker) *-ku-* (object concord of noun class 15) *-bon-* (verb root = "see") *-a* (verbal ending). We do not foresee to enable our system to analyse such input data, however, linguistic verbs consisting of several morphemes should, in principle, be analysed so that users can receive the data on the items related to their query. Users interested in stems on the other hand, should also be able to query those and get the data on all full forms containing a particular stem.

Concerning the disjunctively written Sotho languages, there are other challenges: The copulative of Northern Sotho, for example, consists of one or several, disjunctively written morphemes. These morphemes are highly ambiguous and the copulatives generated from them are homographous, too. The many forms cannot all be described in a printed dictionary due to space constraints. However, even in an electronic dictionary, the task of describing all forms might turn out to be too complex. An attempt has been made to extract these forms from corpora by using regular expressions (Faaß and Taljard 2013), however, due to the many homographs, no system to distinguish them could be found. Such rather morpho-syntactic challenges can be related to the issue of accessibility. We therefore do not see the electronic dictionary itself as the best solution, but rather develop connected systems that could, for instance, assist learners in producing the correct form, such as a decision tree-like device (described in Prinsloo, Bothma, Heid and Faaß 2012).

In an electronic dictionary, these analyses of input data, however, belong to access structure (see section 5), not to the data storage itself. One could, therefore, argue that in a lexicographic electronic data collection there is no macrostructure at all.

We place the sense element at the centre of our database, and since we link this sense with one (or more) orthographic forms and with a stem, we enable our system to allow for immediate access to stems of verbs and nouns, for instance, the Northern Sotho verb stem *bona* "[to] see", but also to full forms such as the Zulu address *sobonana* "see you (again)". Therefore, in terms of

orthographic forms, we foresee simplex and complex words which are both related to sense elements.

The change of focus is exemplified in the following two figures. Figure 1 illustrates a possible entry describing the English verb "[to] see" and its Zulu counterpart "[uku]bona" in a traditional lexicographic database where the lemma is the central element, and is linked to two senses, each extended with an example. The two translation equivalents are linked with each other.

In Figure 2, the same data is viewed from the perspective of our proposed model where English and Zulu data are entered independently, similar to Figure 1. The relational table "is_translation_of" informs that sense 1 and sense 3 are translation equivalents. Note that in Figure 1, the metaphorical sense of "[to] see"/"[uku]bona" was described in each language in the element "sense 2". In the new model, such a sense description does not appear as such. Instead, a literal sense description of "[to] understand"/"[uku]qonda" is included together with an example ("I understand what you mean."). The literal senses 1/2 ("[to] see"/"[to] understand") and senses 3/4 ("[uku]bona"/"[uku]qonda") are then linked with each other by items in the table "is_synonym_to" (see section 3.3). In this table, we learn that the synonymy is metaphorical and we also see the respective example sentences ("I see what you mean" | "Ngiyabona ukuthi uthini").

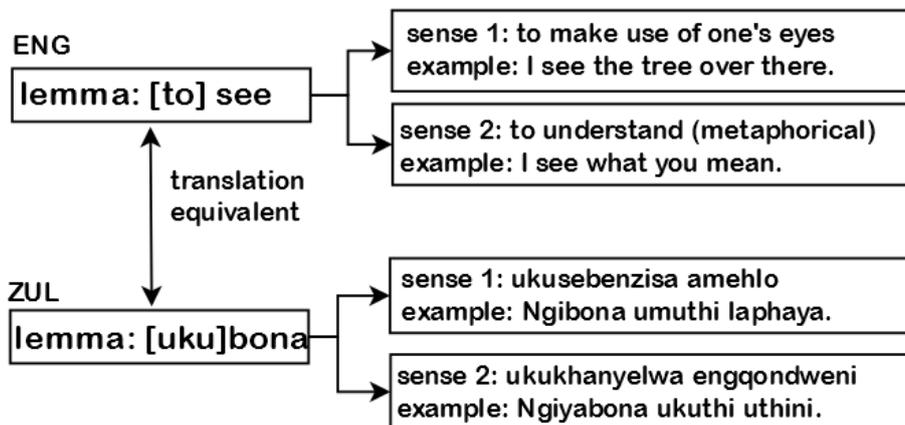


Figure 1: Illustration of the traditional data model: focus on the lemma

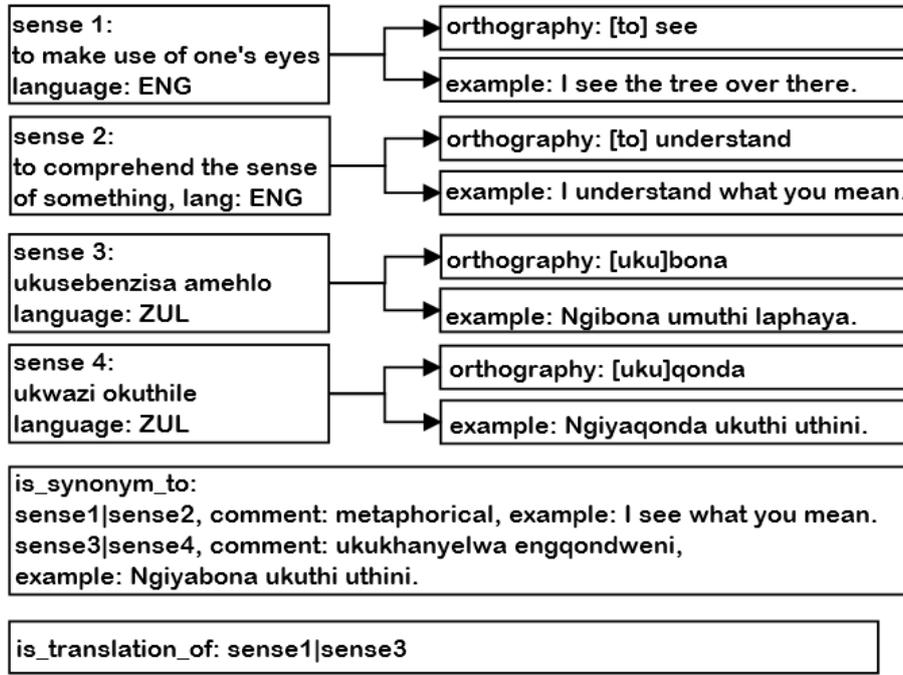


Figure 2: Illustration of the proposed data model: focus on the sense description

3.2 Microstructural items

We began with a general list of items which are usually part of the microstructural items in any dictionary, such as the lemma sign, its paraphrases of meaning, etcetera. For each of these items, we decided whether we require them for our database. Afterwards, we added all items that usually appear in the respective African language dictionaries that we are concerned with. We subsequently categorised the items, which we currently foresee: we generally differentiate between the categories "descriptions", "morpho-syntax", "phonetics", "etymology", "valency", "examples" and "idioms". Each of the tables representing these categories contains its microstructural items. As described above, we need to differentiate between data items to be filled for stems (the ones that are not identical with full forms) and data items to be filled for full forms. Table 1 shows the items foreseen, irrespective of the language they belong to. For the African languages, we add information on whether the item is described for full forms, for stems or for function words.

<i>Data category</i>	<i>Item giving the</i>	<i>Gloss</i>	<i>Full form y/n</i>	<i>Stem y/n</i>	<i>Function words y/n</i>
sense	short paraphrase of meaning	one or several brief semantic descriptions of what the item means (in comparison to other lemmas)	yes	optional	optional (for adv. prefixes/particles)
	paraphrase of meaning	one or several long semantic descriptions of what the item means (in comparison to other lemmas)	yes	optional	optional (for adv. prefixes/particles)
	source	source of brief or long semantic description	yes	optional	optional
	style marker	groups of humans who use the item (e.g. woman only, closed set)	yes	no	no
	subject area	what kind of subject does the item refer to (closed set)	yes	no	no
morpho-syntax	class	free unit (syntactically free: simple or complex, or idiom), or clitic (proclitic, mesoclitic or enclitic), or bound unit (stem or affix) see also Spohr (2012: 69)	yes	yes	yes
	abbreviation	an abbreviation of the item	yes	no	no
	degree of comparison	describes the degree of the item (the positive is not marked): comparative, superlative	yes	no	no
	gender	gender of the item (closed set)	optional (only persons)	no	no
	part-of-speech	morpho-syntactic classification (closed set)	yes	no	yes
orthography	grapheme lemma sign	one or several surface forms of the item in question	yes	no	yes
phonetics	pronunciation	pronunciation of the item (we have yet not decided on the format of this item, we however foresee using IPA)	yes	no	yes
	stressed syllable(s)	indicate which syllables of the item are stressed	yes	no	yes
	syllable division	result of syllabification	yes	no	yes
etymology	long description	long description of the etymology of the item (not a word formation issue)	yes	no	optional
	short description	brief description of the etymology of the item (not a word formation issue)	yes	no	optional
	source	source of long/short description	yes	no	optional
examples	example phrase	a phrase in which the item typically occurs in (to demonstrate the use of the item)	yes	no	yes

	example sentence	a sentence in which the item typically occurs in (to demonstrate the use of the item)	yes	no	yes
	source	source of example phrase or sentence	yes	no	yes
idioms	fixed expression	one or more example(s) of a (partially) idiomatic phrase (e.g. "kick the bucket" in English) which the item occurs	yes	no	yes
	idiom	one or more example(s) of an idiom ("der Krug geht so lange zum Brunnen bis er bricht") that the item occurs in	yes	no	no
	frequency of occurrence	how often does the idiomatic sense of the item occur in a corpus which — if possible — should be near representative; eventually we will use relative frequencies to abstract from the size of the corpus	yes	no	no
	source	source of fixed expression or idiom	yes	no	yes

Table 1: Microstructural items for all languages contained in the database

The items we need for the African languages only are listed in Table 2.

<i>Data category</i>	<i>Item giving the</i>	<i>Gloss</i>	<i>Full form y/n</i>	<i>Stem y/n</i>	<i>Function words y/n</i>
morpho-syntax	government	this item is in general used for items that govern the case in which another must appear (some German prepositions, for example, require their argument to appear in the dative). Concerning the African languages, we make use of this item to describe the influence of a conjunction on the verbal moods to follow it	no	no	yes
	noun class	the noun classes of the Bantu languages replace number/gender classes of other languages	yes	no	optional
phonetics	tone marker	high or low tone might lead to a difference in meaning	yes	yes	yes

Table 2: Microstructural items contained for the African languages only

Lastly, Table 3 contains the items only used for Afrikaans or English, respectively. We do not claim the tables to be comprehensive, other items might be added at a later stage.

Languages	Category	Item	Gloss
Afrikaans	morpho-syntax	attributive marker	marks the adjective taking a derivational "-e" when being used attributively
Afrikaans/ English	morpho-syntax	case	case that a verb refers to (closed set)
Afrikaans/ English	morpho-syntax	inflection	person, number and gender that a verb refers to (closed set)

Table 3: Microstructural items necessary for non-African languages only

3.3 Relational tables

While and after the data items are stored in the database with their respective descriptive items, additional tables describing the relations between them will be defined. In addition to the usual morpho-syntactic relations (e.g. "is-plural-of"), semantic relations are described too (e.g. "is-near-synonym-of"). So far, we do not foresee adding WordNet data. However, this is possible from a technical perspective, since the development of a prototype African Wordnet (AWN), which currently includes four languages, is an on-going project (Griesel and Bosch 2014). The resource has been developed by translating Common Base Concepts (CBC) from English and currently holds roughly 42 000 synsets.

To assign translation equivalents, we use the relation "is-translation-of". A rather general relation will be added as well: "is-linked-with" will contain relations between items not described in the others (i.e. miscellaneous kinds of relations that appear not frequent enough to give reason for an own relational table). This last table, however, will contain a data field where the type of relation is explained.

We relate senses of lemmas with the following tables:

- is-diminutive-of (for nominal items only)
- is-plural-of (for nominal items only)
- is-locative-of (for nominal items only)
- is-stem-of (see lemmatisation strategy above)
- is-homonym-of
- is-near-synonym-of
- is-antonym-of
- is-translation-of (relates items of different languages to each other)
- is-contained-in-example-sentence
- is-contained-in-fixed-expression

- is-contained-in-idiom
- has-morpho-syntax (relates a specific id of a type of morpho-syntactic item to one sense)
- has-phonetics
- has-valency (relates a specific id of a type of valency to one sense of an item taking arguments)
- is-linked-with

For space reasons, we describe only two of the tables in the following sections.

3.3.1 "has-morpho-syntax"

In any typical dictionary, the microstructure contains information on morphology and syntax of a lemma. Such information is repetitive not only for parts of speech appearing several times, but also for their morphological properties. Plural morphemes of English, for example the "-s" appearing in nouns like "type – types", "house – houses", must only be described once in our model. We foresee to fill a table called "morpho-syntax" with all the appearing categories (e.g. *noun*, -s). Each of the categories receives a unique id. In the relational table "has-morpho-syntax", we link the sense descriptions with one or several id(s) of morpho-syntactic categories that apply to them.

3.3.2 "has-valency"

Concerning the valency (or "valence", as described by Spohr 2012: 86f) of a lexicographic item, a similar situation occurs: one type of valency, for example "*verb, taking no object*" can be linked with several words ([to] sit⁵, [to] walk, etc.). We handle the situation in the same way as the "has-morpho-syntax"-table described above. A unique id is assigned to each valency type and sense descriptions are then related to the ids that apply to them.

Some relations between items will be added manually. For this purpose and the purpose of checking and correcting the data that will be inserted automatically (see section 6.4), the database will offer a maintainer interface.

4. Design and implementation method

In this section, we compile the items described above and define a basic lexicographic model where each category represents one table of the database (DB), see Figure 3 which, due to space constraints, does not show all of the items. In our model, we tentatively define relations between items, however, keeping them open for future changes by storing them into separate tables. In MySQL,

each item is identified via an "id"-data element (e.g. "sense-id" identifying one specific paraphrase of meaning). Such identifiers are marked as "primary key", which means that each may only appear once in the respective table. In the model shown in Figure 3, each of the items contained are to be pre-defined in respect of their type, "int" stands for integer, "varchar" for any kind of character. Lastly, "link" means that a URL will be entered.

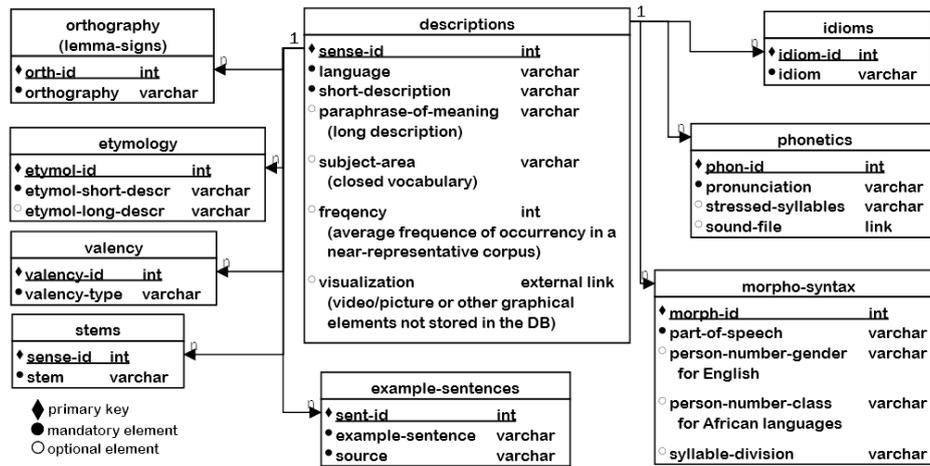


Figure 3: The basic database model showing tentative relations between data items

In an SQL database, relations between items of tables are to be described, which reflect dependencies between items (we can also define item as "hierarchies", as it is done in XML or in object-oriented database systems). A paraphrase of meaning, for example, should directly be related to one or several example sentences, similar to an integrated microstructure. The relation between those items is, therefore, 1:n where "n" stands for any integer number greater than zero. For example, the relation between the items "sense-id" of the table "descriptions" and "sent-id" of the table "example sentences" could be defined as "1:n". However, it could very well be the case that we could use one example sentence several times, by assigning several lemmas (or rather senses of those words) to it, therefore, we do not enforce the 1:n relation by directly linking items (e.g. foreign keys), but rather implement the word sense/example sentence relation by assigning a unique key to each of those items in the respective tables and by adding a separate table linking those ids to each other, see Figure 4.

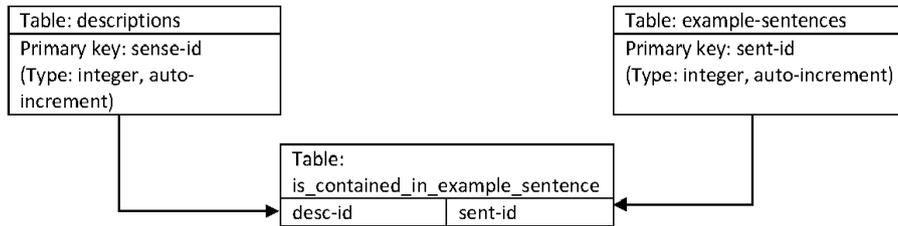


Figure 4: Adding relations between word-sense and example-sentence

The positive aspect of such an implementation is its openness towards a re-
definition of relations between items; a negative aspect might be that such
tables lead to a slow query processing of the database. Therefore, in our second
phase of implementation (i.e. after the available data will have been stored in
the database), we will automatically generate additional tables each containing
all relevant data for one of the dictionaries. Users will have access to each one
of these tables with one mouse click and one or several query words.

5. Data presentation: access structure

Bergenholtz and Gouws (2010: 103) maintain that "of critical importance in a
user-driven lexicographic approach is the need to ensure that the target users
of a specific dictionary gain unimpeded access to the data they need in order to
achieve an optimal retrieval of information". Such accessibility is typically
ensured by the access structure of any given dictionary. We adhere to the defi-
nition of Wiegand and Beer (2013: 111), who define accessibility as follows:
"The term 'data accessibility' refers to the access willingness and thereby to the
possibility to look up textual and illustrative lexicographical data; it is given
because the data are in the access domain of an access structure. A distinction is
made between the external and the internal data accessibility".

In printed dictionaries, the first step is determined by the knowledge a
user has of the specific dictionary. A user could embark on either the full or a
shortened outer access process, reaching the desired lemma via a rapid access
structure, for instance; thumb index markers or alphabet letters, or by merely
guessing where the relevant item will be and then following the running heads
until the desired page has been reached. Going down the lemmata, the desired
guiding item can then be found — the item, where the inner access route com-
mences. In e-dictionaries, a single word or multi-word string is typically typed
into the search box and this will immediately guide the user to the required
lemma sign without bringing any other outer access items into play. Other
systems offer a rapid access structure in the form of a list of clickable lemma

signs of which the user can select the required one. It is also possible to offer both, as described by Bothma and Gouws (2013).

The selection and the order of appearance of the data items both depend on several factors: (1) The type of dictionary; a bilingual dictionary will require a translation equivalent to appear, while a monolingual will not. (2) The part of speech of the lemma; some parts of speech need to be displayed with valency information, for others, valency plays no role. (3) The access route; the first resulting screen of a query will display only few items, from there, the user may click respective boxes on the screen to get more data (e.g. etymological information, idioms or example sentences). For each of the microstructural items above we need to define when it will appear on the screen (given that an orthographic form was entered as a query and this form was found in the database). Table 4 shows these decisions for several of the microstructural items above when a general monolingual dictionary is compiled; due to space constraints, not all assignments can be shown.

	<i>Only specific part-of-speech</i>	<i>Access route: first</i>	<i>Access route: more info</i>	<i>Solely on demand</i>
short paraphrase of meaning	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
paraphrase of meaning	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
source of paraphrase	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
style marker	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
subject area	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
class	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
abbreviation	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
degree of comparison	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
noun class	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tone marker	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
etymological short/long description	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
example sentence	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
fixed expression	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Table 4: Examples of microstructural items being assigned to specific use situations

In section 5 above, we mentioned that for each of the foreseen dictionaries we will generate one table in the database containing all the necessary data. Table 4 above shows their elements for the planned monolingual dictionary of Northern Sotho.

5.1 External links

From a technical perspective, the database is planned to be connected, *inter alia*, with a morphological analyser. This is essential especially for the African languages that are written conjunctively; a user may enter, for instance, the orthographic word *abazukukhombisa* "they will not show it" — without knowing that this expression consists of a number of morphemes: *a-* (negative morpheme), *-ba-* (subject concord class 2), *-zu-* (future tense negative morpheme), *-ku-* (object concord class 15), *-khomb-* (verb root), *-is-* (causative extension), *-a* (verbal ending). Whenever such a query word is not found as a lemma by the database, this morphological analysis will be executed in order to deliver the linguistic units and their parts of speech which will be queried automatically by the system. The user will then see the results for each of the parts presented by the system and can select the items he or she is interested in to get further information displayed.

On the other hand, a user might enter a stem of a word; in this case, we will use the morphological analyser as generator and will generate full form words which could be queried in the database. It is foreseen to then suggest this list to the user, in order for the user to subsequently choose the ones he or she wants to know more about. Concerning productive purposes, we also foresee (user-activated) connections with the decision-tree system developed in the framework of the SeLA project (e.g. described in Prinsloo, Bothma, Heid and Faaß 2012). Another option will be to access corpus data, however, only maintainers will be allowed to see the whole of the data, as one cannot assume that all corpus data would be usable for exemplifying the meaning of a word (see section 6). The maintainers then will be able to choose example phrases or sentences to be added to the database.

6. Resources to be added to the database

It would be virtually impossible to fill such a database from scratch — corpora are scarce and the ones that do exist lack a description of their contents and are, therefore, not feasible for an automated retrieval of dictionary contents. However, there are some resources that we can indeed utilise for a start, as described below.

6.1 Available resources for the project

Language data for Northern Sotho is currently available in the form of a printed dictionary (Ziervogel and Mokgokong 1985), which was scanned⁶ into electronic format by means of Optical Character Recognition (OCR) and transformed at least partially to a structured data collection (Kebbe 2013). We also use a MySQL database containing about 600 full Zulu forms and their English

translation equivalents, generated in the SeLA sub-project on a Zulu dictionary of possessive constructions (Bosch and Faaß 2014). Lastly, we also have access to a file containing several thousand Xhosa nominal stems, information on the noun classes they appear in and their translations into English.

6.2 Other possible resources

In South Africa, the co-ordination of language resources is still in its infancy stages, however, the function of the newly established Language Resource Management Agency (RMA) is to develop and host reusable text and speech resources, and to manage and distribute these from one central point. Currently, relevant resources available are Annotated Text Corpora for all official languages of South Africa annotated with lemma, part of speech and morphological analyses. Initial versions of core technologies, namely lemmatisers, part of speech taggers and morphological decomposers are available as open source modules and could, therefore, be used for the annotation of text corpora of the various Bantu languages, although (Eiselen and Puttkammer 2014: 3702) point out that "there is still a lot of room for improvement, especially for lemmatisation and morphological decomposition".

6.3 Adding resources to the database

Despite several corpora for the African language that we are permitted to use for the purpose of, for instance, checking corpus frequencies of occurrences⁷ to be added (manually) to the database at a later stage, we were also able to get access to a scanned dictionary (Ziervogel and Mokgokong 1985). Unfortunately, the files we received were in word format, and all items were in the same font, so it was impossible to automatically identify item types by their format. Judith Kebbe, a student of information science at University of Hildesheim worked out an automated method to identify item types by their position in the dictionary article (Kebbe 2013) and wrote Perl scripts extracting those items, based on the descriptions of Faaß, Ramagoshi and Sebolela (2009). Her work resulted in structured, machine-readable data covering about half of the entries of the dictionary. As it turned out, however, the microstructure of this dictionary is not structured consistently; when trying to extract translation equivalents, especially, the automated method often failed. Another problem is described by Kosch (2013: 204) who points out the mixed lemmatisation approach of this dictionary, whereby a word approach is applied to nouns with irregular or non-overt class prefixes, although the overriding approach in the dictionary is stem based. The user is then given a cross-reference to the relevant stem. Examples are nouns such as *mmuši*, "ruler" and *pono* "vision", which are lemmatised as words and not as stems. According to the stem-based approach, the lemmatisation of the two nouns would have presented as *buši* and *bono*,

derived from the verb stems *-buša* "rule" and *-bona* "see" respectively.

Kebbe extracted several thousand links between dictionary entries, but only few dictionary entries describing translation equivalents. Hence, these data will be loaded into our database to cater for monolingual Northern Sotho only, mainly to test relational tables such as for instance "is-linked-with".

Bosch and Faaß (2014) populated a MySQL database with about 600 Zulu nouns and about 900 English translation equivalents, there is also information on their classes and numbers stored in this database. We will transfer these data to the SeLA database as well.

Bilingual Xhosa–English data was made available to us in .xls format. Here, not surface forms but several thousand noun stems, the classes they appear in, class prefixes, and English translation equivalents are contained. By way of shell scripts, we will generate full forms and fill the database with the respective data.

With the available resources, we cannot fill the sense descriptions in most cases; therefore, we will have to add them manually. During the import of the data, we plan to use English translations to have these mandatory fields filled, but these will have to be replaced manually with monolingual sense descriptions. As our team will not have the manpower to fill all of the foreseen database items, we plan to send out calls to the public, trying to find volunteers, as soon as the graphical user interfaces have been completed. For our aim to compile prototypical dictionaries, we consider the available data to be sufficient.

6.4 An example: monolingual Northern Sotho data

This article describes a lexicographic model which is still awaiting implementation. While implementing it, we might find errors or inconsistencies that will force us to change the model. Therefore, at this stage, we can only describe data that was examined during the development of the model. We chose the dictionary of Ziervogel and Mokgokong (1985) that contains several thousand noun stems with additional information. One of the dictionary entries contains data on the noun stem *mente*:

```
MENTE, -/di- (mêntê) munt (waar geld geslaan word) // mint (where money  
is coined)
```

Kebbe (2012:34) generated the following, machine-readable data from of this entry:

```
<entry>  
  lemma:MENTE  
  cppl:di  
  <translation>  
    <Afrikaans>munt (waar geld geslaan word)</Afrikaans>  
    <English>mint (where money is coined)</English>  
  </translation>  
</entry>
```

With these data, we cannot provide a Northern Sotho sense description to fill the mandatory item "short description" in the "descriptions"-table of our database. In a first attempt, we hence foresee to write scripts that make use of the English translation. The scripts however add the note "TO-BE-TRANSLATED-INTO-NSO" as an indication for the manual reworking which is foreseen at a later stage. The "language"-field can be filled automatically because we know that this is NSO data. Optional elements (as shown in Figure 3) are not filled:

1. Table "descriptions":
sense-id: 1,
language: NSO,
short-description: TO-BE-TRANSLATED-TO-NSO: mint (where money is coined),
paraphrase-of-meaning: empty field,
subject-area: empty field,
frequency: empty field,
visualization: empty field.

Next, we process the information on morpho-syntax: cppl stands for "class prefix plural" which is an indication that this new database entry describes a noun. As this prefix is *di* and as no singular prefix is given, the scripts can assume automatically that the noun is of class 9 which means that its orthographic form of the singular is identical to the stem (*mente*). Therefore the singular orthographic form is *mente* and the plural form *dimente*. Since this dictionary uses diacritics to indicate tone, we also learn about the high tone on the second vowel. The scripts can hence fill several tables:

2. "morpho-syntax":
morph-id: 1, part-of-speech: noun, person-number-class: 03-sg-09,
morph-id: 2, part-of-speech: noun, person-number-class: 03-pl-10.
3. "stems"
stem-id: 1,
stem: mente.
4. "orthography"
orth-id: 1, orthography: mente,
orth-id: 2; orthography: dimente.
5. "phonetics"
phone-id: 1,
pronunciation: mêtê,
stressed-syllables: empty field,
sound-file: empty field.

Lastly, the scripts will fill the necessary relational tables creating links between the items.

1. "is-plural-of":
orth-id:2 | orth_id:1.

2. "is-stem-of":
stem-id:1 | orth-id:1,
stem-id:1 | orth-id:2.
3. "has-morpho-syntax":
sense-id:1 | morph-id:1.
4. "has-phonetics":
orth-id:1 | phone-id:1.

7. Summary and future work

This article describes the design of a lexicographic data model which will be implemented with MySQL, resulting in a database capable of storing lexicographic data of several of the official languages of South Africa. We aim at compiling several prototypical dictionaries from there: a monolingual Northern Sotho dictionary, a bilingual Xhosa–English general language dictionary and a bilingual English–Zulu learners' dictionary. We have compiled lists of necessary microstructural elements and have decided to put the sense description at the centre, the "lemma" being just a realisation of the sense, in other words its surface form.

We have collected a number of resources, which will be loaded onto the database semi-automatically. At this stage, it is foreseen that all missing data items will require manual adding due to the lack of available resources. It is well known that the development of resources for African languages is often of a fragmented nature — the resources tend to be small, only usable for restricted purposes and, therefore, excluding connection with other resources. We, therefore, intend to investigate collaborative approaches and technologies for the accumulation and creation of data to ensure the continued filling of this lexicographic database (cf. Benjamin 2014).

8. Endnotes

1. The term "African languages" refers to languages belonging to the Bantu language family. Both terms are used in this article.
2. Francopoulo, G. (Ed.). 2013. *LMF Lexical Markup Framework*. London: Wiley-ISTE. ISBN: 978-1-84821-430-9.
3. SeLA is supported by the "Deutscher Akademischer AuslandsDienst", DAAD in their programme "Welcome to Africa", see also www.uni-hildesheim.de/iwist-cl/projects/sela/.
4. phpMyAdmin is available from www.phpmyadmin.net
5. For ease of understanding, we make use of the orthographic forms in our example. In the data model foreseen, we will however relate sense descriptions with the "has-valency" ids.
6. We gained permission from the publisher to use the scanned dictionary at least for our current research purposes and hope that we will be allowed to use the resulting data for the prototype of our dictionaries.

7. We are very grateful to Prof. Prinsloo for allowing us to use his African Languages corpora (see De Schryver and Prinsloo 2000).

9. Acknowledgement

This article has been written in the framework of the project *Scientific eLexicography for Africa* (SeLA) working towards the development of electronic dictionaries and terminology databases for the (South) African Bantu Languages. The project is funded by the German ministry for education and research (BMBF, Funding number 55516493) in the Framework of the DAAD programme "Welcome to Africa" (see also <http://www.uni-hildesheim.de/iwist-cl/projects/sela/>).

10. Bibliography

- Benjamin, M.** 2014. Collaboration in the Production of a Massively Multilingual Lexicon. *Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era. LREC, 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland, May 26-31, 2014*: 211-215.
- Bergenholtz, H. and R.H. Gouws.** 2010. A New Perspective on the Access Process. *Hermes* 44: 103-127.
- Bergenholtz, I. and H. Bergenholtz.** 2013. One Database, Four Monofunctional Dictionaries. *Hermes — Journal of Language and Communication in Business* 50: 119-125.
- Bosch, S.E., L. Pretorius and J. Jones.** 2007. Towards Machine-readable Lexicons for South African Bantu Languages. *Nordic Journal of African Studies* 16(2): 131-145.
- Bosch, S.E. and G. Faaß.** 2014. Towards an Integrated e-Dictionary Application: The Case of an English to Zulu Dictionary of Possessives. *Proceedings of the 16th Euralex Conference. Proceedings of the Sixteenth EURALEX International Congress, EURALEX 2014, Bolzano/Bozen, Italy, July 15-19, 2014*: 739-747. Bolzano/Bozen: Eurac.
- Bothma, T.J.D. and R.H. Gouws.** 2013. Mapping Indicators in Print Dictionaries to Structural Descriptions in e-Dictionaries. Paper read at the *eLex 2013 Conference, Tallin, Estonia, October 17-19, 2013*. Tallinn, Estonia: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000. The Compilation of Electronic Corpora, with Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.
- Eiselen, R. and M.J. Puttkammer.** 2014. Developing Text Resources for Ten South African Languages. *Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era. LREC, 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland, May 26-31, 2014*: 3698-3703.
- Faaß, G., R. Ramagoshi and F. Sebolela.** 2009. Updating the Setswana Monolingual School Dictionary, *Thanodi ya Setswana ya Dikole: An Experimental Study*. Otlogetswe, Thabelo J. (Ed.). 2009. *MLA Kgasa: A Pioneer Setswana Lexicographer*: 146-172. Cape Town: Centre for Advanced Studies of African Society (CASAS), book series 64.
- Faaß, G., S.E. Bosch and E. Taljard.** 2012. Towards a Part-of-Speech Ontology: Encoding Mor-

- phemic Units of Two South African Bantu Languages. *Nordic Journal of African Studies* 21(3): 118-140.
- Faaß, G. and E. Taljard.** 2013. Automatic Detection of Copulatives in Northern Sotho Corpora. (online) *Book of Abstracts on the 5th International Conference on BANTU Languages, Paris, September 2012*.
- Gouws, R.H. and D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds).** 2013. *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/New York: De Gruyter.
- Griesel, M. and S.E. Bosch.** 2014. Taking Stock of the African Wordnet Project: 5 Years of Development. Orav, H., C. Fellbaum and P. Vossen (Eds). 2014. *Proceedings of the Seventh Global WordNet Conference (GWC 2014), Tartu, Estonia, 25–29 January 2014*: 148-153. Tartu: GWA. http://gwc2014.ut.ee/proceedings_of_GWC_2014.pdf.
- Heid, U.** 2012. SeLA — A New Project on Electronic Lexicography. *Lexicographica* 28(1): 437-440.
- ISO 24613:2008.** *Language Resource Management*. Edited by International Organization for Standardization, Geneva, 2008.
- Kebbe, J.** 2013. *Creating a Pre-dictionary Fact Base from a Printed Trilingual Dictionary*. Hildesheim: Bachelor project (BA International Information Science), University of Hildesheim, Institute for Information Science and Natural Language Processing (available on request).
- Kosch, I.** 2013. Expectation Levels in Dictionary Consultation and Compilation. *Lexikos* 23: 201-208.
- L'Homme, M.-C.** 2012. Adding Syntactico-semantic Information to Specialized Dictionaries: An Application of the FrameNet Methodology. *Lexicographica* 28: 233-252.
- Prinsloo, D.J., U. Heid, T. Bothma and G. Faaß.** 2012. Devices for Information Presentation in Electronic Dictionaries. *Lexikos* 22: 290-320.
- Prinsloo, D.J.** 2010. Internet Dictionaries for African Languages. *Lexicographica* 26: 183-194.
- RMA.** [accessed: 2014]. Language Resources Management Agency. <http://rma.nwu.ac.za>.
- Scholze-Stubenrecht, W.** 2013. The World Wide Web as a Resource for Lexicography. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds). 2013: 1365-1374.
- Spohr, D.** 2012. *Towards a Multifunctional Lexical Resource: Design and Implementation of a Graph-based Lexicon Model*. *Lexicographica*. Series Maior 141. Supplementary Volume. Berlin/Boston: Walter de Gruyter.
- Wiegand, H.E. and S. Beer.** 2013. Access Structures in Printed Dictionaries. Gouws, R.H., U. Heid, W. Schweickard and H.E. Wiegand (Eds). 2013: 110-149.
- Ziervogel, D. and P.C. Mokgokong.** 1985. *Groot Noord-Sotho-Woordeboek: Noord-Sotho, Afrikaans/Engels*. Pretoria: J.L. Van Schaik.

Buitetekste in 'n elektroniese gebaretaalwoordeboek*

Hanelle Fourie Blair, *Buro van die Woordeboek van die Afrikaanse Taal, Stellenbosch, Suid-Afrika en Departement Afrikaans en Nederlands, Universiteit Stellenbosch, Stellenbosch, Suid-Afrika*
(hfourie@sun.ac.za)

Opsomming: Die voorstelle wat in hierdie artikel oor buitetekste in 'n elektroniese gebaretaalwoordeboek gemaak word dien as deel van 'n volledige konsepmodel wat spesifiek vir die De la Bat Skool vir Dowes in Worcester, Suid-Afrika, voorgestel word (Fourie 2013). Die doel van die voorgestelde woordeboek is om die leerlinge van die De la Bat Skool elektroniese toegang tot geskrewe skoolwoordelyste te gee, met gebare en voorbeelde in gebaretaal wat in videoformaat vertoon word — dus in die vorm van 'n elektroniese woordeboek. 'n Geskrewe **woordelys** voldoen glad nie in hierdie geval aan die behoeftes van skoolleerders nie en 'n **woordeboek** wat vir die Dowe teikengebruikers by die skool ontwerp is, sal van veel groter waarde wees in die leerders se omgang met beide gebaretaal sowel as die geskrewe taal.

Die artikel bespreek dataverspreiding en die gebruik van 'n raamstruktuur, die gebruik van buitetekste en verskaf ook voorbeelde van buitetekste in bestaande gebaretaalwoordeboeke, hetsy elektronies of nie. Daarna word voorstelle vir buitetekste in die konteks van die model gemaak met illustrerende voorbeelde.

Sleutelwoorde: AFRIKAANS, BUITETEKSTE, DATAVERSPREIDING, DOWE TEIKEN-GEBRUIKERS, ELEKTRONIESE WOORDEBOEKE, GEBARETAAL, GRONDSLAGFASE, LEKSIKOGRAFIE, LEKSIKOGRAFIESE FUNKSIES, MINIGRAMMATIKA, RAAMSTRUKTUUR, WOORDEBOEK

Abstract: Outer Texts in an Electronic Sign Language Dictionary. The suggestions made in this article about outer texts in an electronic sign language dictionary serve as part of a complete concept model suggested specifically for the De la Bat School for the Deaf in Worcester, South Africa (Fourie 2013). The purpose of the proposed dictionary is to give the learners of the De la Bat School electronic access to written school word lists, with signs and examples in sign language displayed in video format — therefore in the form of an electronic dictionary. A written word list does not meet the needs of school learners in this case and a dictionary that is designed

* Hierdie artikel is 'n uittreksel uit 'n Ph.D.-proefskrif, *'n Leksikografiese model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans*, wat in Maart 2013 deur die Universiteit Stellenbosch, Stellenbosch, Suid-Afrika, aanvaar is. Dit is ook aangebied as lesing by die Negentiende Jaarlikse Internasionale Konferensie van die African Association for Lexicography (AFRILEX), georganiseer deur die Navorsingseenheid vir Tale en Literatuur in the SA Konteks, Noordwes-Universiteit, Potchefstroomkampus, Potchefstroom, Suid-Afrika, 1–3 Julie 2014.

for the Deaf target users at the school will be of far greater value in the learners' communication in both sign language and the written language.

The article discusses data distribution and the use of a frame structure, the use of outer texts and also supplies examples of outer texts in existing sign language dictionaries, whether electronic or not. Following that suggestions for outer texts in the context of the model are made with illustrating examples.

Keywords: AFRIKAANS, DATA DISTRIBUTION, DEAF TARGET USERS, DICTIONARY, ELECTRONIC DICTIONARIES, FRAME STRUCTURE, FOUNDATION PHASE, LEXICOGRAPHIC FUNCTIONS, LEXICOGRAPHY, MINI GRAMMAR, OUTER TEXTS, SIGN LANGUAGE

Inleiding

Hierdie artikel ondersoek die moontlikhede ten opsigte van buitetekste wat in 'n elektroniese tweetalige woordeboek van Suid-Afrikaanse Gebaretaal (SASL¹) en Afrikaans vir die grondslagfase gebruik sou kon word. Die voorstelle wat gemaak word dien as deel van 'n volledige konsepmodel wat spesifiek vir die De la Bat Skool vir Dowes in Worcester voorgestel word (Fourie 2013). Die doel van die voorgestelde woordeboek is om die leerlinge van die De la Bat Skool elektroniese toegang tot hul skoolwoordelyste (soos in die volgende paragraaf beskryf) te gee, met gebare en voorbeelde in gebaretaal wat in videoformaat vertoon word — dus in die vorm van 'n **elektroniese woordeboek**.

Leerlinge is reeds vertrou met 'n woordelys vir elke graad of fase op skool, en hierdie lys is tematies gegroepeer. Dieselfde lys word dwarsdeur die grondslagfase gebruik, en die leerders merk die leksikale items wat hulle reeds geleer het, daarop af. Op hierdie manier weet die onderwyser in die volgende klas/graad watter woordeskat nog geleer moet word en watter bloot hersien kan word. Daar is ook ruimte op die lys vir leerders om hul eie woordeskat by te voeg en nuwe inskrywings te maak. Die nadeel van die geskrewe lys is dat dit leerders slegs items uit die kernwoordeskat in die *geskrewe* taal, nl. Afrikaans, bied. Dit bevat ook geen illustrasies of enige ander visuele leidrade wat as geheuehulp t.o.v. die *gebare* van hierdie kernwoordeskat (of die denotata waarop die woorde en gebare dui) kan dien nie. Dit verskaf dus slegs 'n lys van die *inhoud* van die geskrewe woordeskat wat in die verskillende leerareas van die grondslagfase baasgeraak moet word en bevat glad nie voorbeeldmateriaal van die gebruik van hierdie woordeskat nie. Dit is duidelik dat 'n geskrewe **woordelys** — veral die woordelyste in hul huidige vorm — glad nie aan die behoeftes van skoolleerders voldoen nie en dat 'n **woordeboek** wat vir die Dowe teikengebruikers by die skool ontwerp is, van veel groter waarde sal wees in die leerders se omgang met beide gebaretaal sowel as die geskrewe taal.

Woordeboeke, ook elektroniese woordeboeke, word beskou as tekssamestellings en draers van verskillende tekssoorte (Gouws 2001: 60). Buitetekste

kan beskou word as addisionele tekste wat die sentrale lemmalys aanvul en steeds deel het aan die leksikografiese funksies van die betrokke woordeboek en wat tot die gebruiker se behoeftes t.o.v. kennis en kommunikasie bydra. In die geval van 'n elektroniese woordeboek verteenwoordig buitetekste dus enige materiaal wat ook buite die databasis staan of as subafdeling bestaan van die databasis waaruit soekresultate gegenereer word. In 'n elektroniese woordeboek kan buitetekste saam gegroepeer en op 'n spyskaart gelys word of hulle kan d.m.v. opwiptekste ("pop-up texts") of soortgelyke skakels van binne die mikrostruktuur bereik word.

1. Dataverspreiding

Die funksionaliteit van 'n beplande woordeboek strek verder as die perke van die sentrale lemmalys: beide die voor- en agtertekste kan data en teks bevat wat 'n funksionele rol in die aanbieding van leksikografiese inligting het (Gouws en Prinsloo 2005: 58). 'n Raamstruktuur bied meer opsies aan die leksikograaf wanneer leksikografiese aanbod beplan word — buitetekste help nie net die gebruiker om suksesvolle woordeboekraadplegingsprosedures en die beste verkryging van inligting te verseker nie, maar speel ook 'n belangrike rol in die dataverspreidingstruktuur van die woordeboek deur die leksikograaf toe te laat om leksikografiese data in meer as een teks aan te bied (Gouws en Prinsloo 2005: 58).

Die gebruik van buitetekste vereis 'n besluit oor die inligting wat daarin aangebied moet word en die verhouding tussen die buitetekste en die sentrale lys. Twee hooftypes dataverspreidingstrukture kan geïdentifiseer word, nl. 'n **eenvoudige dataverspreidingstruktuur** en 'n **uitgebreide dataverspreidingstruktuur**: waar die sentrale lys die enigste teiken vir dataverspreiding is, vertoon die woordeboek 'n eenvoudige dataverspreidingstruktuur; waar buitetekste of dele van buitetekste gebruik word om data as deel van die prosedure van dataverspreiding aan te bied, vertoon die woordeboek 'n uitgebreide dataverspreidingstruktuur (cf. Bergenholtz, Tarp en Wiegand 1999: 1779).

Die sentrale of hoofteks word in 'n woordeboek met 'n raamstruktuur deur die buitetekste omraam, nl. die voor- en agtertekste. Die voortekste-afdeling bevat gewoonlik onder meer:

1. die verpligte gebruikersgids,
2. 'n beskrywing van een of meer aspekte van die woordeboek se objektaal.

Die agtertekste-afdeling is opsioneel en bevat gewoonlik addisionele tekste, byvoorbeeld met getalle, plekname, nasionaliteite, tale, ens. (cf. Hausmann en Wiegand 1989: 331). Geen buitetekste behoort egter ooit as blote kosmetiese byvoegsel gesien te word nie aangesien hulle 'n belangrike rol in die oordrag van data in woordeboeke behoort te speel (Gouws 2009a: 67) en die leksikograaf ook toelaat om kategorieë inskrywings in te sluit wat nie gewoonlik in die

betrokke soort woordeboek se artikels verskyn nie (Gouws en Prinsloo 2005: 60).

Woordeboekinleidings of gebruikersgidse word gewoonlik beskou as hoogs waardevolle metaleksikografiese tekste wanneer hulle deur 'n (bedrewe of vaardige) gebruiker benut word. Hulle bevat inligting oor onder meer die inhoud en strukture van die betrokke woordeboek (Mavoungou 2002: 105). Al-Ajmi (2001) het ses fokusareas geïdentifiseer wat gebruik kan word om die inleidings van woordeboeke krities te evalueer, naamlik:

1. Spesifisering van die beoogde gebruikersgroep,
2. Aanduiding van die redes waarom die woordeboek saamgestel is,
3. Verwysing na die bronne van die woordeboek,
4. Oorsig van nuwe verskynsels in die inleiding,
5. Spesifisering van die aantal lemmas (of woorde),
6. Praktiese leiding t.o.v. die gebruik van die woordeboek.

In die konteks van hierdie navorsing wat gerig is op 'n spesifieke gebaretaalwoordeboek, behels sommige van hierdie punte die volgende: die beoogde gebruikersgroep is geïdentifiseer as die Dowe² leerders wat onderrig by die De la Bat Skool vir Doves in Worcester ontvang. Horende gebruikers, soos die ouers, vriende, familie en selfs onderwysers van hierdie leerders, word as die sekondêre teikengroep beskou, aangesien dié woordeboek nie in die eerste plek daarop gemik is om horendes te help of te leer om gebaretaal te gebruik nie. Hierdie woordeboek hoop om vir die eerste keer in Suid-Afrika weg te beweeg van 'n benadering waar die fokus op horende-Dowe interaksie val, en die uitgangspunt is eerder 'n klemverskuiwing waar die fokus op gebaretaal self val, in hierdie geval SASL.

2. Die toepassing of gebruik van buitetekste

Bergenholtz, Tarp en Wiegand (1999: 1776-1778) onderskei tussen **geïntegreerde** en **niegeïntegreerde** buitetekste, wat dit vir die leksikograaf moontlik maak om 'n stel buitetekste te ontwerp met of sonder 'n direkte skakel met die werklike doel van die spesifieke woordeboek. Tekste soos die lys medewerkers, die historiese ontwikkeling van die taal, en dies meer is tipiese voorbeelde van niegeïntegreerde buitetekste aangesien hulle gewoonlik nie 'n direkte skakel met die sentrale lys van 'n woordeboek of sy ware doel toon nie. In teenstelling hiermee is die voorbeeld van die *Longman Dictionary of Contemporary English* (LDOCE) (Procter 1978) se tabel met diereenname en o.a. hul manlike, vroulike en verkleinvorme, soos in voorbeeld 4 aangegee, 'n voorbeeld van 'n geïntegreerde buitetekste aangesien dit die gebruiker die moontlikheid gee om direk by hierdie inligting (wat ook in die sentrale lys verskyn) uit te kom en dit in tabelvorm te beskou sonder om die sentrale teks te raadpleeg.

In die geval van data wat in tabelvorm as buitetekste gegroep word en wat ook in ('n tematiese kategorie van) die sentrale lys verskyn, soos syfers in die lys wat hieronder in paragraaf 4.5 aangegee word, skep 'n multimedia- of elektroniese formaat die moontlikheid om ook 'n hiperskakel na die inskrywing vir die spesifieke item in die sentrale lys te verskaf; die lemma word van 'n volledige bewerking voorsien wat inligting soos die handvorm, die plek in die gebare ruimte, die geskrewe ekwivalent van die gebaar, woordsoort, 'n lynskets van die gebaar, 'n video van die gebaar, 'n geskrewe glos van 'n voorbeeld van die gebaar in 'n volledige sin sowel as 'n video van die gegewe voorbeeld en die korrekte vertaling van die gebaretaaluiting in die geskrewe taal insluit. 'n Voorbeeld van hierdie tipe toepassing van 'n buitetekste verskyn in die *Online Dictionary of New Zealand Sign Language* (NZSL Online) (McKee 2011): die gebruiker kan in die lys hoofgetalle (ingedeel onder die hoofopskrif "Numbers" wat in 'n navigeringsbalk geplaas is wat altyd sigbaar is) op die skakel wat saam met die lynskets van die getal verskyn, klik:



Figuur 1: Navigeringsbalk in die NZSL Online waarop die skakel na 'n 'buitetekste' bestaande uit 'n oorsigtelike groepering van die gebare vir getalle gevind word



Figuur 2: Lynskets van die gebaar vir die getal 9 wat in die NZSL Online op die bladsy onder die hoofopskrif "Numbers" verskyn

(Die tussenstap, nl. die bladsy direk onder die skakel "Numbers" waar al die gebare oorsigtelik saamgegroepeer is, is in die bostaande voorbeelde oorgeslaan om ruimte te spaar.)

Deur op die skakel (9) wat by die gebaar verskyn, te klik, word die gebruiker herlei na die bladsy waar die volledige artikel van die gebaar aangebied word:

 	<p>nine 9 numeral</p> <p>Add to Vocab Sheet</p>	
	 <p>Play in slow motion</p>	
	<p>Usage Examples</p>  <p>Play in slow motion</p>	<p>The boy invited <i>nine</i> friends to his birthday party.</p> <p>boy invite nine friend come birthday party</p> <p>Play in slow motion</p>

Figuur 3: Volledige artikel van die gebaar vir die getal 9 in die NZSL Online

Dit kan hier genoem word dat dit belangrik is dat hiperskakels verkieslik altyd soos "tipiese" hiperskakels moet lyk (dus in blou en onderstreep, soos in die skakel vir (9)), aangesien die deursneegebruiker van die internet a) reeds verwag dat 'n hiperskakel so sal lyk en b) ook sal verstaan dat teks wat op hierdie manier gemerk is, 'n skakel is wat na 'n ander bladsy sal lei.

Soos reeds genoem, is buitetekste meer as net kosmetiese versierings (Gouws 2009a) of selfs byvoegsels wat na willekeur bygevoeg of weggelaat word, maar hulle behoort 'n ondersteunende rol tot die hoofartikels van die woordeboek t.o.v. die oordrag van inligting te speel. 'n Woordeboek behoort ook 'n transtekstuele benadering tot leksikografiese funksies te hê wat mee-

bring dat die funksies wat in die sentrale lys nagestreef word, ook deur buite-tekste bevredig moet word.

2.1 Tabele en lysse in geskrewe woordeboeke

'n Voorbeeld van 'n buiteteks wat aan goeie leksikografiese beginsels voldoen en die ware doel van die woordeboek ondersteun deur die gebruiker te help om inligting wat in die woordeboek is (makliker) te bekom, en wat ook 'n kognitiewe funksie het, is die dieretabel in die LDOCE (Procter 1978: 1298):

ANIMAL	MALE	FEMALE	YOUNG	GROUP NOUN	Noise — all verb and noun: some are also interjections
cat	tom(cat)	queen	kitten		purr miaow <i>interj</i>
cattle (<i>pl.</i>)	bull	cow	calf	herd	bellow (of bull) moo (of cow) <i>interj</i>
chicken	cock	hen	chick	brood (of young)	
dog	dog	bitch	pup(py)		bark whine growl bowbow <i>interj</i>
dolphin, porpoise, whale	bull	cow	calf	school	
donkey					heehaw <i>interj</i> bray
duck	drake	duck	duckling		quack <i>interj</i>
deer (<i>pl.</i> deer)	buck stag	doe hind (esp. of reddeer)		herd	
fish				shoal	
fox	dog	vixen	cub		bark
goat	billy (goat)	nanny (goat)	kid	herd	bleat
goose (<i>pl.</i> geese)	gander	goose	gosling	flock gaggle	hiss honk
horse	stallion	mare	foal	herd (esp. in the wild)	neigh whinny
lion	lion	lioness	cub	pride	roar

pig	boar	sow	piglet	herd	grunt oink <i>interj</i>
rabbit	buck	doe			
seal	bull	cow	pup	colony	bark
sheep	ram	ewe	lamb	flock	bleat baa <i>interj</i>

Figuur 4: Dieretabel in die LDOCE (Procter 1978: 1298)

Tabelle of lyste wat tipies in die buitetekste van woordeboeke vir jong gebruikers, soos die *Oxford First Bilingual Dictionary* (OFBD) (Hiles en Redelinghuijs 2010) en die *Longman Grondslagfasewoordeboek Afrikaans/Engels* (LGFW) (Gouws et al. 2010), gebruik word, sluit in lyste van die alfabet, die kalender (dae van die week en maande van die jaar), seisoene, syfers/getalle, teenoorgesteldes, kleure en geometriese vorme. (Sien Addendum A–C.)

Die gebruik van kleurvolle prentjies om konsepte soos teenoorgesteldes en die gebruik van voorsetsels te illustreer, is nie alleen prakties vanuit 'n leksikografiese oogpunt nie, maar ook aantreklik en prettig vanuit die jong gebruiker se oogpunt. Deur aan jong gebruikers te demonstreer dat dit ook lekker kan wees om iets in die woordeboek te soek en/of iets daaruit te leer met die hulp van kleurrike sketse, kan 'n woordeboekkultuur (mettertyd) gekultiveer word by 'n gebruikersgroep wat tot dusver (in die spesifieke geval van die voorgestelde woordeboek) afgeskeep is.

3. Voorbeelde van buitetekste in bestaande gebaretaalwoordeboeke

'n Oorsig van die voor- en agtertekste wat in enkele bestaande gebaretaalwoordeboeke aangebied word, gee 'n goeie idee van die tipe buitetekste wat vir die gebruiker van 'n gebaretaalwoordeboek van nut kan wees en dus ook vir gebruik in die beplande woordeboek *oorweeg* kan word.

Stokoe et al. (1976) se *Dictionary of American Sign Language on linguistic principles* (voorts DASL) beskik oor 'n omvattende beskrywing van ASL en die hantering daarvan in die woordeboek (in 'n inleidende voorteks), die sintaksis van ASL (bylaag A — al die bylaes word as agtertekste aangebied), naamgebare ("name signs", dus die gebaar wat aan 'n spesifieke persoon toegeken word om 'n voorstelling van sy of haar naam te wees en wat dikwels deur Dowes gekies word om 'n spesifieke eienskap van die persoon te beklemtoon) en die getalstelsel (bylaag B), die taalkundige gemeenskap van ASL (bylaag C), gebaretaaldialekte (bylaag D) sowel as 'n alfabetiese indeks van Engelse woorde (en frases soos "I didn't mean that", "turn one's stomach", "going back and forth", "hanging out laundry"). Geen eksplisiete verduideliking van dié Engelse woorde word gegee nie, maar dit wil voorkom asof dit die Engelse ekwivalente vir die gebare in die sentrale lys is — die woordeboek is georden volgens gebare (volgens Stokoe se notering van die gebare volgens handvorm en ander parameters) en nie volgens geskrewe woorde nie. Die Engelse indeks

dien dus as 'n kruisverwysing na items in die sentrale lys.

Dit is interessant om op te merk dat die indeks ook eiename bevat, soos "Lincoln, Abraham", "Benson, Elizabeth E.", en plekname, soos "Hartford, Conn." en "Phoenix, Arizona". Die artikel vir "Benson, Elizabeth E." dui aan dat dit 'n naamgebaar is en lyk soos volg:

Elizabeth E. Benson, Dean of Women, Gallaudet College. See Appendix B (Stokoe et al. 1976: 185).

Die artikel bevat dus 'n kruisverwysing na die bylaag oor naamgebare, aangesien dit 'n belangrike komponent van ASL (en ander gebaretaal) is wat in meer besonderhede in die bylaag bespreek word.

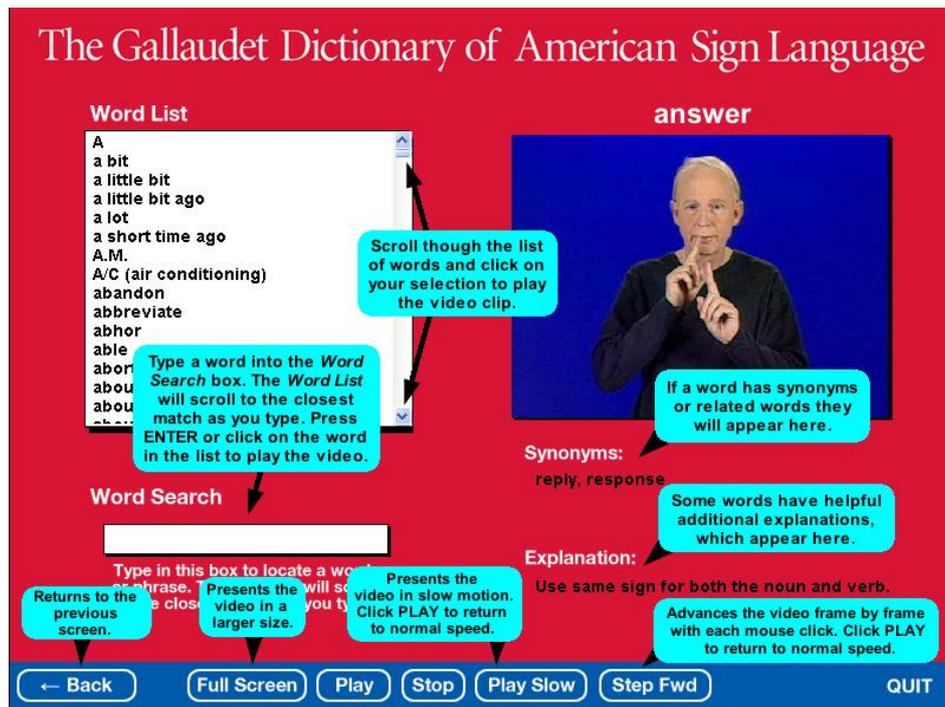
Die *Dictionary of British Sign Language/English* (DBSL) (Brien 1992) beskik oor 'n **inleidende voorteks** en **gebruikersgids** wat in inhoud en aanbieding baie ooreenkomste toon met die woordeboeke van gesproke/geskrewe tale. Dit bevat o.a. 'n afdeling oor die geskiedenis van die DBSL, 'n oorsig van gebaretaalleksikografie oor die algemeen, 'n kort opsomming van die inhoud van die woordeboek, 'n omvattende gebruikersgids onder die opskrif "Locating information in the Dictionary" wat van voorbeeldbladsye uit die woordeboek voorsien is, 'n uiteensetting van die notasiesisteme soos wat dit in die DBSL gebruik is, 'n uiteensetting van die niemanuele eienskappe, 'n produksiegids (wat gebruikersleiding gee oor die korrekte gebruik van die gebareparameters handvorm en oriëntasie (gids een), plek, rangskikking van die hande en kontak (gids twee), beweging (gids drie) en niemanuele eienskappe. (Sien Addendum D-F.)

Daar is ook 'n baie handige deel met "Questions and answers" wat voorbeelde verskaf van die maniere waarop die woordeboek gebruik kan word, bv. wat om te doen in die geval van 'n gebruiker wat a) vlot is in BSL en b) oor 'n mate van kennis van Engels beskik, wat c) besig is om 'n brief in Engels te skryf en wil d) uitvind of daar ander Engelse woorde met dieselfde betekenis (dus sinonieme) is vir 'n woord wat dikwels in die teks gebruik word (dus **teksproduksie** vir die Dowe gebruiker van die geskrewe taal). 'n Ander scenario sluit in 'n gebruiker wat a) vlot is in BSL en b) oor 'n mate van kennis van Engels beskik, wat c) besig is om 'n boek te lees en d) op soek is na ander betekenis vir 'n spesifieke woord waarvan een betekenis aan die gebruiker bekend is (dus **teksresepsie** vir die Dowe gebruiker van die geskrewe taal). In terme van horende gebruikers word daar o.a. voorsiening gemaak vir scenario's waarin 'n vlot spreker van Engels met 'n mate van kennis van BSL 'n gebaar in gesprekke teëgekomp het maar nie weet hoe om dit te produseer nie (dus **teksproduksie** vir die horende gebruiker van gebaretaal); 'n vlot spreker van Engels met 'n mate van kennis van BSL wat 'n video in BSL het waarin daar 'n gebaar is wat hy/sy nie verstaan nie (dus **teksresepsie** vir die horende gebruiker van gebaretaal); 'n vlot spreker van Engels met 'n mate van kennis van BSL wat op soek is na 'n gebaar om 'n spesifieke betekenis in 'n gebaretaaluitlegging oor te dra (dus **teksproduksie** vir die horende gebruiker van gebaretaal). 'n Uiteensetting van volledige instruksies vir elke scenario wat geskets word, word verskaf en bied

aan die gebruiker 'n stap-vir-stap-gids oor presies hoe om die woordeboek te gebruik om in spesifieke behoeftes te voorsien.

In 'n aparte afdeling getiteld "The Visual World of British Sign Language: an introduction" bied Brennan (1992: 1-133) 'n omvattende beskrywing van die Dowe gemeenskap en veral die taal, insluitend die vasgestelde en produktiewe leksikon, klassifiseerders en die grammatika (woordklasse in BSL, werkwoordverbuiging, ontkenning, die breër grammatikale konteks, ens.). Hierdie beskrywing kwalifiseer dalk nie as 'n "minigrammatika" nie, maar is wel bruikbaar en nodig, veral in die lig van die wanopvattinge en verkeerde of gebrekkige kennis wat daar oor die algemeen oor gebaretaal bestaan.

Die *Gallaudet Dictionary of American Sign Language* (Valli 2005) (voortaan GDASL) op CD-ROM bevat 'n hulpknoppie as toegang tot 'n tipe buitetekst, aangesien die gebruiker na die volgende beeld gelei word wanneer die knoppie gedruk word:



Figuur 5: Hulpbladsy met instruksies soos gevind op die GDASL CD-ROM (Valli 2005)

Hierdie bladsy is nuttig omdat dit 'n beskrywende etiket aan alles op die skerm gee en verduidelik wat die funksie of betekenis daarvan is. Die enigste knoppie op die Help-bladsy wat aktief is en geklik kan word, is die "Back"-knoppie, wat die gebruiker terugneem na die oorspronklike, soekbare gebruikerskoppelvlak.

Die GDASL in boekvorm sluit benewens 'n **algemene inleiding** met 'n oorsig en geskiedenis van die Amerikaans Dowe gemeenskap, 'n bespreking van ASL en die eienskappe en struktuur daarvan (insluitend die gebareparameters en niemanuele eienskappe of tekens), 'n afdeling oor die uitbreiding van die leksikon deur middel van vingerspelling, leenwoorde uit ander (gebare)tale en verandering oor 'n verloop van tyd, variasie en kontak tussen tale ook 'n aparte afdeling oor **klassifiseerders** in ASL in sowel as 'n lys **Engelse woorde** wat gewoonlik in ASL **gevingerspel** word, hetsy volledig of afgekort:

English Words That Are Usually Fingerspelled

The following words are usually fingerspelled as complete words or as abbreviations.

air = A-I-R	nail = N-A-I-L
apartment = A-P-T	nap = N-A-P
bank = B-A-N-K	news = N-E-W-S
bar = B-A-R	oil = O-I-L
barbecue = B-B-Q	ounce = O-Z
cab = C-A-B	pan = P-A-N
car = C-A-R	prescription = R-X
ferry = F-E-R-R-Y	quarterback = Q-B
handicapped = H-C	racquetball = R-B
headquarters = H-Q	rice = R-I-C-E
heel = H-E-E-L	roll = R-O-L-L (bread)
high school = H-S	rug = R-U-G
ice = I-C-E	social security number = S-S + NUMBER
junior = J-R	tablespoon = T-B-S (measure)
liberal = L-I-B-E-R-A-L	teaspoon = T-S-P (measure)
map = M-A-P	Total Communication = T-C
menu = M-E-N-U	tip = T-I-P
meter = M-E-T-E-R	toe = T-O-E
mud = M-U-D	yard = Y-A-R-D
multiple sclerosis = M-S	

Figuur 6: Lys woorde wat gewoonlik in ASL gevingerspel word (Valli 2005: xii)

Hierdie lys in die GDASL is gerig op die bereiking van die werklike doel van die woordeboek, nl. die resepsie en produksie van ASL. Die gebare wat in die lys genoem word, word egter nie in die gedrukte teks herhaal nie en slegs

enkeles verskyn wel op die CD-ROM (die hele GDASL bestaan uit 'n gedrukte woordeboek en 'n multimedia woordeboek op CD-ROM wat soekbaar is volgens die geskrewe woorde maar wat ook video-opnames van die gebare bevat). Daar word 'n onderskeid getref tussen woorde wat gewoonweg gevingerspel word en woorde wat oorspronklik uit die gesproke/geskrewe taal geleen is maar wat **geleksikaliseer** is in ASL, bv. #JOB en #BACK (die hutsteken # word by geleksikaliseerde vingerspelgebare gebruik) — die oorgang tussen die onderskeie letters word so vloeiend dat sommige van die letters uitgelaat word, bv. in die gebaar #JOB word slegs die J en B gevingerspel, en in #BACK slegs die B, A en K. Dit wil egter voorkom asof die GDASL hierdie onderskeid nie konsekwent toepas nie, aangesien die vingerspelgebaar BUS (waarin elke letter B, U en S gevingerspel word) as die geleksikaliseerde inskrywing #BUS opgeneem is.

Die idee behoort egter nooit te wees om elke moontlike buitetekste in die voorgestelde skoolwoordeboek of enige ander (skool)woordeboek te gebruik nie, aangesien dit na alle waarskynlikheid slegs na informasiedood ("information death", cf. Bergenholtz et al. 2009; Bergenholtz en Gouws 2010) sal lei. Die *Finger Talk South African Sign Language (SASL) Dictionary* (Howard 2008) (voortaan FTSASLD) bied 'n kompakte versameling buitetekste van die noodsaaklikste inligting aan, wat moontlik as 'n riglyn vir die insluiting van buitetekste in die voorgestelde woordeboek gebruik kan word: benewens 'n inhoudsbladsy en 'n kort inleiding (waarin dit lyk asof die woordeboek hoofsaaklik op die ouers en vriende van die Dowe leerders by die Fulton Skool vir Dowe gemik is, aangesien die woordeboek onder die naam en embleem van die skool verskyn) bevat die voortekste ook 'n afdeling oor die **gebareparameters** en **nienmanuele eienskappe** of tekens, die **vingerspelalfabet**, **maande** van die jaar en **getalle** 0 tot 20. Die sentrale lys van die FTSASLD is ingedeel in tematiese kategorieë soos Kommunikasie, Plaasdiere, Wilde Diere, Klere, Liggaamsdele asook Werkwoorde, Vraagwoorde, Voornaamwoorde, Voorsetsels en Teenoorgesteldes — hierdie kategorieë stem tot 'n groot mate ooreen met die tematiese kategorieë wat op De la Bat se klaswoordelyste verskyn, wat dit laat voorkom asof die teikengebruikers wel die Dowe leerders by die skool (ook) is. Ongelukkig word die teikengebruiker(s) nie pertinent geïdentifiseer nie.

Na die sentrale lys volg 'n **alfabetiese indeks** van al die woorde wat in die woordeboek verskyn tesame met die bladsynommer waar hulle gesien kan word. Dit maak van die woordeboek 'n politoeganklike bron. Daar is ook 'n bladsy met "Useful Contacts" waarop die adresse en kontakbesonderhede van o.a. DEAFSA, die Dowe Vereniging van KwaZulu-Natal, 'n gebaretaalinstrukteur en die KwaZulu-Natal Blinde en Dowe Vereniging aangegee word. Hierop volg weer twee bladsye met 'n kort geskrewe teks oor die Fulton skool sowel as verskeie foto's, en heel laaste 'n bladsy waarop notas gemaak kan word en waarop kortliks inligting omtrent die outeur verskyn.

Die agtien (18) handvorme wat in die FTSASLD gelys word, is veel minder as die ses-en-sestig (66) handvorme wat in totaal in die *Dictionary of South-*

ern African Signs (Penn 1992: 12) gelys word. Dit lyk egter nie asof die FTSASLD enige gebare bevat wat uit handvorme bestaan wat nie in die voorteks gelys is nie; dit veroorsaak dus geen probleme binne die konteks van die woordeboek en sy gebruik nie. Op soortgelyke wyse hoef die voorgestelde skoolwoordeboek nie 'n lys of indeks van elke *moontlike* handvorm in SASL te bevat nie, maar die lys en indeks kan saamgestel word op grond van die handvorme wat in die opname van die gebare wat gebruik gaan word, gevind word.

4. Voorstelle vir buitetekste

4.1 Beskrywing van die brontaal en doeltaal

Die primêre teikengebruiker van die voorgestelde woordeboek is 'n Dowe leerder wat nie noodwendig baie goeie taalvaardigheid in die tweede (geskrewe) taal het nie. Die situasie word egter verder gekompliseer deurdat die tipiese Dowe leerder in sommige gevalle ook nie 'n baie gesofistikeerde of uitdruklike kennis van gebaretaal het nie, aangesien die oorgrote meerderheid Dowe kinders nie gebaretaal van hul ouers leer nie en dikwels eers in 'n later stadium daaraan blootgestel word, soos by 'n residensiële skool. Daarby is die onderrig en bestudering van gebaretaal as 'n formele skoolvak nog 'n nuutjie en word die gebaretaalkurrikulum ten tyde van hierdie skrywe steeds in fases in die verskillende grade in die skool bekendgestel.

Die sekondêre teikengebruiker is 'n horende leerder van gebaretaal, veral iemand soos 'n familielid wat self met die basiese woordeskat vertrouwd wil raak sodat hulle (makliker) met die Dowe leerder kan kommunikeer.

Beide gebaretaal en Afrikaans tree in die voorgestelde model as bewerkingstaal én bewerkte taal op wat meebring dat die voorgestelde woordeboek volkome tweetalig is en dus vereis dat lemmabewerkings volledig en parallel in beide tale aangebied word.

Daar moet dus aanvaar word dat beide die geskrewe taal en gebaretaal deeglik behandel sal moet word en dat die fokus eweredig tussen die twee tale gebalanseer sal moet word. Hierdie ewewig in die hantering van albei tale moet ook in die buitetekste van die voorgestelde woordeboek weerspieël word. Die voorgestelde woordeboek verteenwoordig die eerste geval waar gebaretaal self (ook) as brontaal in Suid-Afrika aangebied word. Die buitetekst van die voorgestelde woordeboek sal dus nie net 'n minigrammatika van die geskrewe taal moet aanbied nie, maar ook 'n grammatikale oorsig van gebaretaal. Hierdie benadering sal in ooreenstemming wees met die **ware doel** van die woordeboek, nl. om die Dowe gebruiker (asook die horende gebruiker) te help om die toepaslike inligting in albei tale so gou moontlik op te spoor.

'n Fonologiese, morfologiese en selfs sintaktiese analise van beide tale sal dus waarde tot die voorgestelde woordeboek toevoeg. Die woordeboek sal nie net 'n leksikografiese verwysingbronn wees nie maar ook 'n taalkundige hulpmiddel (ook in die resepsie en produksie van die geskrewe taal).

Daar bestaan geen twyfel daaroor dat die primêre teikengebruikers en hul naslaanvaardighede die struktuur van die woordeboek behoort te bepaal nie. Daar is nie tans 'n woordeboek met SASL as brontaal nie en daar is nog baie debat oor die standaardisering van die taal. Volgens Gallardo (1980: 61) behoort daar nie in die geval van 'n taal wat nog in die proses van standaardisering is prioriteit aan die samestelling van 'n eentalige woordeboek gegee te word nie, maar eerder aan die samestelling van 'n tweetalige of selfs veeltalige woordeboek. Die voorgestelde woordeboek voldoen dus aan hierdie aanbeveling.

4.2 Instruksies aan die gebruiker

Gegewe die feit dat die primêre teikengebruiker geen ervaring van woordeboeke en 'n beperkte woordeboekkultuur het — veral waar dit nie moontlik is om 'n gebaar op te soek sonder die hulp van 'n geskrewe taal nie — sal die buitetekste vir besonder deeglike instruksies en opleiding aan die gebruiker voorsiening moet maak (benewens die duidelike leiding wat ook in die klas aan die gebruikers gegee sal moet word). Voorbeelde uit al drie dele van die teks, nl. 'n tematiese kategorie en 'n tipiese artikel of subveld (of 'n kort deeltrajek) daaruit, 'n voorbeeld van die manier waarop gebare volgens handvorm gesoek word sowel as die manier waarop 'n woord as soekmetode optree, moet ingesluit en duidelik geïllustreer en van aanwysings in beide die geskrewe taal sowel as gebaretaal voorsien word om aan die gebruiker te wys hoe om inligting nie net vanuit die (drie) soekmetodes nie, maar ook vanuit die mikrostruktuur binne elke artikel te ontgin. Hierdie instruksies kan d.m.v. 'n videotutorial aangebied word. 'n Goeie voorbeeld van 'n omvattende stap-vir-stap videotutorial in gebaretaal (naas die geskrewe instruksies) kan aanlyn op die NZSL Online-webblad by <http://nzsl.vuw.ac.nz/help> gesien word.

Dit is belangrik om dit eksplisiet of stap-vir-stap aan die gebruiker duidelik te maak hoe en in watter situasies die woordeboek gebruik kan word, en dit kan gedoen word deur sekere scenario's te skets wat dien as voorbeelde van die tipiese funksies waarop die woordeboek gemik is. **Teksproduksie** in die geskrewe taal kan verduidelik word aan die hand van die Dowe leerder wat weet wat 'n gebaar beteken maar nie weet wat die geskrewe ekwivalent vir daardie gebaar is nie. **Teksresepsie** in die geskrewe taal kan weer geïllustreer word as 'n Dowe leerder wat 'n woord in 'n teks lees maar nie seker is wat die woord beteken nie. Op 'n soortgelyke manier behoort die gebruikershandleiding die sekondêre gebruikersgroep, nl. horende leerders van SASL, se behoeftes oor die produksie en resepsie van SASL te hanteer en te verduidelik hoe die woordeboek die gebruiker in daardie aspekte ondersteun, bv. in gevalle waar 'n horende leerder a) nie weet watter gebaar om in 'n sekere situasie te gebruik nie (en dus 'n soektog via 'n geskrewe woord na 'n gebaar sal volg), of b) nie weet wat 'n gebaar in 'n uiting beteken nie (en dus 'n soektog via 'n gebaar na 'n geskrewe woord sal volg).

Die instruksies kan op die **Tuis-bladsy** aangebied word, maar 'n skakel

daarheen kan via 'n hulp-knoppie of -skakel op elke elektroniese 'bladsy' van die gebruikerskoppelvlak ("user interface") ingesluit word sodat dit altyd teenwoordig en bereikbaar is. Veral internet- en ook die meeste ander rekenaar-gebruikers is daaraan gewoond dat daar 'n hulpfunksie of -knoppie in die betrokke program is en hulle soek intuïtief daarvoor **op die bladsy waar hulle is**. Vir 'n voorbeeld van 'n hulpbladsy, sien die bespreking van die *Gallaudet Dictionary of American Sign Language* (Valli 2005) se hulpbladsy in die paragraaf 3 hierbo.

Dit behoort duidelik aan die gebruiker gestel te word dat die inhoud van die woordeboek gebaseer is op die temas wat volgens die betrokke leerplanne in die klas behandel word, waarvoor die enigste hulpmiddel tot dusver die geskrewe woordelyste was — deur die nuwe woordeboek op hierdie manier binne die konteks van die leerder se daaglikse situasie te plaas, behoort dit vir die leerder duidelik te wees *hoekom* daar nou 'n woordeboek is en *waarvoor* hy/sy dit kan gebruik.

Die **keuse** en **doel** van die ander tipe buitetekste wat gekies en ingesluit word, soos tabelle van diere en hul geluide, die trappe van vergelyking, teenoorgesteldes, voorsetsels, kleure, ens. (sien paragraaf 4.5.) moet ook deeglik geïllustreer en verduidelik word sodat die teikengebruiker weet a) **wat** in hierdie tekste aangebied word, b) **waar** om dié tekste te vind en c) **hoe, hoekom** en **wanneer** hulle die tekste kan raadpleeg om hul gebruik van die woordeboek aan te vul en uit te brei. Net soos in die geval van 'n hulp-ikoon wat op die Tuis-bladsy en ook elke ander bladsy teenwoordig is om die gebruiker na die gebruikersaanwysings te neem, kan 'n ikoon of ikone wat die gebruiker na die buitetekste lei, op dieselfde manier aangewend word.

4.3 Minigrammatika van die geskrewe taal

Soos reeds genoem, is dit nie die doel om hier te fokus op presies hoe die inhoud van die grammatika-aanbod in die woordeboek sal lyk nie. Die tipe onderwerpe wat ter sprake kan kom, word hier bespreek. Die spesifieke gebruikersbehoefte sal telkens bepalend wees vir die aard en omvang van die bespreking. Afhangend van die spesifieke geskrewe taal waarmee die Dowe leerder in die klaskamer te make het, sal die onderstaande beskrywings op Afrikaans of Engels van toepassing wees.

4.3.1 Fonologie

Aangesien die doel van die woordeboek nie is om Dowe gebruikers te help om die geskrewe taal te *praat* nie — spraakontwikkeling vind by die skool op grond van leerders se individuele vermoë plaas — sal dit nie nodig wees om enige uitdruklike verduidelikings oor die uitspraak van die gesproke taal in die inleiding in te sluit nie, en ook nie enige fonetiese transkripsies van woorde nie.

4.3.2 Morfologie en sintaksis

Leerkragte moet geraadpleeg word om te bepaal watter morfeme die belangrikste is (of as die belangrikste in die onderrigssituasie geag word) en met watter morfeme die leerders tipies die meeste probleme ondervind, bv. **toekommende en verlede tyd, meervoud, verkleining**, ens. Op hierdie manier kan aandag gegee word aan die toepaslike morfologiese reëls wat in die klassituasie gebruik word en kan hierdie reëls met voorbeelde van die praktiese toepassing daarvan toegelig word.

Die begrip **woordsoorte** kan aan die hand van voorbeelde verduidelik word om leerders aan die verskille tussen die vernaamste woordsoorte soos naamwoorde, byvoeglike naamwoorde, werkwoorde, bywoorde, voorsetsels, en dergelike, te herinner.

Afrikaanse (of Engelse) **woordvolgorde** behoort ook verduidelik en geïllustreer te word, veral aangesien dié volgorde dikwels verskil van die orde van gebare in 'n gebare-uiting. Hier sal die insig en oordeel van die leerkragte weereens 'n belangrike rol speel in die keuse van sintaktiese aspekte wat behandel moet word asook gepaste voorbeeldmateriaal wat die beste geskik sal wees om vir die leerders van waarde te wees.

Die gebruik van 'n woordeboeksterne verwysing kan oorweeg word indien daar toepaslike handboeke, klasnotas of dergelike materiaal bestaan waarmee die leerders vertrouwd is en wat hierdie inligting in meer besonderhede bespreek en verduidelik. Hier kan die beginsel van **geïntegreerde woordeboekgebruik** toegepas word. Geïntegreerde woordeboekgebruik impliseer dat dit 'n wesentlike deel van 'n bepaalde leeropdrag is, en voorvereis die beplanning en samestelling van 'n bepaalde woordeboek as deel van 'n groter pakket naslaanbronne (Gouws 2009b: 83). 'n Tipiese voorbeeld van geïntegreerde woordeboekgebruik word gevind wanneer 'n woordeboek gekoppel is aan 'n bepaalde nieleksikografiese bron en die gebruik van die woordeboek onder meer daarop gerig is om die inhoud van die betrokke bron beter te ontsluit (Gouws 2009b: 83). Die klaswoordelys vorm (ook) so 'n nieleksikografiese bron en vorm ook die basis vir die inhoud van die woordeboek by gebrek aan 'n ander korpus. Geïntegreerde woordeboekgebruik sal in hierdie geval nie net tot groter insig in die betrokke leerareas lei nie, maar terselfdertyd ook tot die verwerwing van woordeboekgebruiksvaardighede en die vestiging van 'n woordeboekkultuur (Gouws 2009b: 83). Ander handboeke en/of tekste wat in die skool gebruik word en wat terme bevat wat in die woordeboek verklaar word, kan aangepas word sodat die betrokke terme in dié tekste aangedui of gemerk word. Sodoende sal die gebruikers mettertyd weet en onthou dat die 'pakket' van woordeboek tesame met klasmateriaal voorsiening maak vir leksikografiese bewerking van die betrokke terme — die wisselwerking tussen klasmateriaal en woordeboek deur middel van kruisverwysings verhef die klasmateriaal dan tot woordeboeksterne geïntegreerde buitetekste (Gouws 2009b: 88, 90).

Die inligting wat in hierdie afdelings aangebied word, behoort in beide

die geskrewe vorm (aangesien dit immers die geskrewe taal is wat bespreek word) sowel as 'n video-opname met SASL aangebied word om maksimum toegang en begrip te verseker en om getrou te bly aan die woordeboek se tweetalige aard.

4.4 Minigrammatika van SASL

4.4.1 Fonologie

Die fonemiese elemente³ van SASL, nl. handvorm, handpalmoriëntering, ruimte/plek en beweging, sal in hierdie afdeling behandel word. Niemanuele eienskappe (dus die gesigsuitdrukings wat grammatikale betekenis het) sal ook hier behandel en aangebied word.

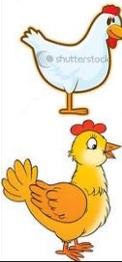
4.4.2 Morfologie en sintaksis

Morfologiese en sintaktiese elemente soos **verbuiging t.o.v. wyse en graad**, **veelvoudigheid** (meervoudsvorming en/of die inkorporering van getal) en **tydsaanduiding** (verlede, toekomstige of teenwoordige tyd), **ruimtelike en persoonsooreenstemming**, die werking en gebruik van **klassifiseerders** as deel van die produktiewe leksikon, **woordsoortelike** indeling en onderskeid, **eenhandige** en **tweehandige** gebare (die dominante en nedodominante hand, tweehandige gebare wat met een gehand gevorm kan word deur die nedodominante hand te laat sak, eenhandige gebare wat met twee hande gevorm kan word deur die dominante hand op die nedodominante hand te dupliseer, ens.), **vingerspelling**, en dies meer, sal in hierdie afdeling bespreek word. 'n Lys klassifiseerderhandvorme kan in tabelvorm aangebied word, asook 'n tabel van die vingerspelalfabet. Al hierdie elemente is reeds vroeër in die artikel verduidelik en geïllustreer en sal nie hier herhaal word nie.

Soos in die geval van die minigrammatika van die geskrewe taal, sal die inligting wat hier aangebied word, in beide SASL sowel as die geskrewe taal verskyn om maksimum toegang te verseker vir sowel die primêre gebruiker, die Dowe leerder, as die sekondêre gebruiker, 'n horende persoon.

4.5 Tabele en lyste

'n Vereenvoudigde weergawe van die LDOCE se dieretabel (figuur 4) met die diere se name en hul geluide soos wat dit in die voorskoolse klaswoordelys verskyn, kan in die buiteteks ingesluit en d.m.v. kruisverwysings in die mikrostruktuur bereik word, bv.:

DIER		DIEREGELUID
	hond	blaf
	kat	miaau
	hoender hen	kekkel
	haan	kraai
	skaap	blêr
	koei	bulk
	perd	runnik
	muis	piep
	leeu	brul

 voël 	sing  fluit 
---	---

Figuur 7: Dierenname en dieregeluide soos wat dit in die voorskoolse klaswoordelys verskyn as een tabel gegroep (die meegaande prentjies is van die internet geneem en is slegs ter illustrasie van die konsep)

Die tabel bied dus toegang tot die woordeboek se inhoud d.m.v. gidselemente (die prente) wat nie tot enige van die lede van die taalpaar van die woordeboek behoort nie, net soos wat die prente die kern van die tematiese soekprosedure vorm (sien Fourie 2013: 303-338 vir 'n bespreking van die databasis en soekprosedures). As gidselemente bied die prente toegang tot die bewerking aan gebruikers van albei die behandelde tale en verhef dit die woordeboek tot iets meer as bloot tweetalig.

Deur die items wat in die skoolwoordelys onder die tematiese opskrif "Plaasdiere" verskyn, eenvoudig in tabelvorm te groepeer, word die direkte verband tussen die manlike en vroulike name vir die onderskeie diere duidelik (bv. **bul** en **koei** vir 'n manlike en vroulike **bees**), sowel as die name wat aan die diere toegeken word wanneer hulle nog klein is (bv. **kalf/kalfie** vir 'n jong **bees**):

	plaasdier		 (Skets plus video van gebaar vir "plaasdier")	
	'n Plaasdier woon op 'n plaas.		 (Video van die definisie in gebaretaal)	
	PLAASDIERE 	MANNETJIE 	WYFIE 	BABA 

	 hoender (mv. hoenders)	  haan (mv. hane)	  hen (mv. henne)	  kuiken (mv. kuikens)/kuikentjie (mv. kuikentjies)
	eend (mv. eende)			eendjie (mv. eendjies)
	skaap (mv. skape)	ram (mv. ramme)	ooi (mv. ooie)	lam (mv. lammers)/lammetjie (mv. lammetjies)
	perd (mv. perde)	hings (mv. hingste)	merrie (mv. merries)	vul (mv. vullens)/vulletjie (mv. vulletjies)
	donkie (mv. donkies)	donkiehings	donkiemerrie	donkievul (mv. donkievullens)/donkievulletjie (mv. donkievulletjies)
	vark (mv. varke)	beer (mv. bere)	sog (mv. sogge/sôe)	varkie
	haas (mv. hase)			hasie
	bok (mv. bokke)	bokram (mv. bokramme)	bokooi (mv. bokooie)	boklam (mv. boklammers)/boklammetjie (mv. boklammetjies)
	muis (mv. muise)			muisie (mv. muisies)
	skilpad (mv. skilpaaie)			skilpadjie (mv. skilpadjies)
	bees (mv. beeste)	bul (mv. bulle)	koei (mv. koeie)	kalf/kalfie (mv. kalfies)
	duif (mv. duiwe)			duifie (mv. duifies)
	hond (mv. honde)	reun (mv. reuns)	teef (mv. tewe)	hondjie (mv. hondjies)

Figuur 8: Inhoud van die tematiese kategorie "Plaasdiere" wat in tabelvorm as buitetekste aangebied kan word

In die gevalle waar daar nie 'n spesifieke woord vir 'n babadier gebruik word nie, soos **eend**, **vark**, **haas**, **muis**, **skilpad**, **duif** en **hond**, kan die gewone ver-

kleinvorm ingevul word om die tabel te voltooi en die verskillende verkleinvorms vas te lê. Hierdie tabelle se geskrewe inhoud kan op sowel teksproduksie as teksresepsie van toepassing en van hulp wees, en dit het ook 'n kognitiewe funksie aangesien dit die opdoen van feitekennis in die sin van die leer van die verband tussen verskillende manlike en vroulike diere, ens., help bewerkstellig.

In die bostaande tabel verteenwoordig die grys plekhouers die sketse van gebare wat die tabelopskrifte vertaal. Slegs die eerste ry van die tabel is van prentjies (wat vanaf die internet bekom is) voorsien om die konsep te illustreer — ander items in die tabel sal ook van toepaslike prentjies of sketse voorsien word. Die prentjies is hiperskakels wat die gebruiker (terug)neem na die onderskeie artikels (dus lei die prentjie van die hen na dieselfde artikel wat die soekresultaat van 'n soektog met die woord "hen" of die gebaar daarvoor sou wees), maar wat ook 'n skakel (sou kon) bevat wat die gebruiker toelaat om die ander items in die tabel na willekeur op te soek, bv. in die geval van "hen":

	<p>hen</p>	 <p>(Skets plus video van gebaar HEN)</p>
	<p>'n Hen is 'n wyfiehoeender. Die hen lê eiers. 'n Hen kekkel.</p>	 <p>(Video van die definisie in gebaretaal asook die voorbeeldsin in gebaretaal)</p>
	<p>Sien nog PLAASDIERE.</p>	 <p>(Video van kruisverwysing) plus skets van gebaar vir "plaasdiere" wat ook 'n hiperskakel is)</p>

Figuur 9: Mikrostrukturele bewerking vir die woord "hen" met 'n kruisverwysing na die buitetekste vir plaasdiere ("[PLAASDIERE](#)") asook 'n kruisverwysing na 'n buitetekste van diere en die geluide wat hulle maak ("[kekkel](#)") (soos uiteengesit in figuur 7 hierbo), wat in die mikrostruktuur van elke referent in die betrokke kategorie teenwoordig sal wees

Toegang tot die buitetekste kan dus in hierdie geval ook deur skakels in die mikrostruktuur verkry word. Daar is effektief drie hiperskakels in die bostaande tabel van die mikrostruktuur: die eerste een, die prentjie van die hen,

lei, soos al die ander prentjieskakels wat in die mikrostruktuur *links* van die geskrewe woord geplaas word, na die tematiese kategorie/prentjiebladsy vir *Plaasdiere* waarop die prentjie van die hen (saam met al die ander referente in die kategorie) gevind kan word; die tweede een, in die voorbeeldsin ("[kekkel](#)"), lei na 'n opwip-buiteteks waarin die tabel van diere en dieregeluide (soos in figuur 7) aangebied word; die derde een, die geskrewe woord [PLAASDIERE](#) asook die skets van die ooreenkomstige gebaar (wat in die bostaande tabelle deur 'n grys 'plekhouer' verteenwoordig word), lei na die buitetekste vir "Plaasdiere" soos wat dit hierbo in figuur 8 aangedui word. Sou die gebruiker "plaasdiere" d.m.v. 'n woord of gebaar opsoek, sal 'n skakel na die buitetekste ook in die soekresultaat voorsien word:

	<p>plaasdiere</p>	  <p>(Skets plus video van gebaar)</p>
	<p>'n Plaasdiere woon op 'n plaas.</p>	 <p>(Video van die definisie in gebaretaal)</p>
	<p>Sien nog PLAASDIERE.</p>	  <p>(Video van kruisverwysing) plus skets van gebaar vir "plaasdiere" wat ook 'n hiperskakel is)</p>

Figuur 10: Die soekresultaat vir "plaasdiere"

Die mikrostruktuur van die bostaande artikel is grootliks dieselfde as die eerste deel van die mikrostruktuur wat vir die buitetekste aangebied word en bo die tabel van diere name staan, behalwe vir die ekstra komponent van die kruisverwysende skakel na die buitetekste self (in die derde ry van die bostaande tabel). Die prentjie van die plaasdiere heel links binne die artikel is die skakel of aanduiding wat die gebruiker (terug)neem na die tematiese kategoriebladsy waar al die inskrywings binne die kategorie aangebied word.

Nog 'n leerarea en die inhoud daarvan wat as buitetekste oorweeg kan word, is die **trappe van vergelyking** van die byvoeglike naamwoorde wat in die woordelyste verskyn. 'n Kruisverwysende skakel a) vanuit die minigrammatika van Afrikaans en b) vanuit die mikrostruktuur van die byvoeglike naamwoord kan die gebruiker d.m.v. 'n ikoon soos  na 'n tabel soos hierdie (ook as opwipitem) lei:

		 -er	 -ste
	vet   Hans is vet . 	vetter   Pieter is vetter (as Hans). 	die vetste   Gert is die vetste (van almal). 
	maer   Tom is maer . 	maerder   Sarie is maerder (as Tom). 	die maerste   Johan is die maerste (van almal). 
	groot	groter	die grootste
	klein	kleiner	die kleinste
	lank	langer	die langste
	kort	korter	die kortste
	hard	harder	die hardste
	sag	sagter	die sagste
	mooi	mooier	die mooiste
	lelik	leliker	die lelikste
	koud	kouer	die koudste
	warm	warmer	die warmste
	vinnig	vinniger	die vinnigste
	stadig	stadiger	die stadigste
	oop	oper	die oopste
	toe	toeër	die toeste
	grof	growwer	die grofste
	glad	gladder	die gladste
	oud	ouer	die oudste
	jonk	jonger	die jongste
	goedkoop	goedkoper	die goedkoopste
	duur	duurder	die duurste
	lig	ligter	die ligste
	swaar	swaarder	die swaarste
	vuil	vuiler	die vuilste
	skoon	skoner	die skoonste
	donker	donkerder	die donkerste
	lig	ligter	die ligste

Figuur 11: 'n Tabel van die trappe van vergelyking, met illustrerende sketse om die konsepte te kontekstualiseer (slegs die eerste twee inskrywings van die tabel is ter illustrasie van sketse en voorbeeldsinne voorsien — die hele tabel as buiteteks sal uiteraard van passende sketse en voorbeeldsinne voorsien word)

Die versoeking moet weerstaan word om te veel aparte buitetekste aan te bied, selfs in die elektroniese medium waar ruimte nie dieselfde rol speel as in die gedrukte medium nie. Die elektroniese omgewing moet so **gebruikersvriendelik** moontlik wees, en dit behels (o.a.) dat die aantal klikke van die muis wat 'n gebruiker moet uitvoer om by die nodige data uit te kom tot die minimum beperk word. 'n Lang lys buitetekste skep twee probleme: a) hoe en waar word hulle op die bladsy of gebruikerskoppelvlak ("user interface") geplaas sodat die gebruiker weet waar om hulle te kry, en b) hoe gaan die gebruiker weet hoe en wanneer om hulle te gebruik? Die elektroniese medium is nie op dieselfde manier lineêr (van voor tot agter, of van agter tot voor) "blaaibaar" as wat 'n boek is nie. 'n Elektroniese woordeboek is nie bloot 'n gedrukte woordeboek wat op 'n skerm en d.m.v. 'n gebruikerskoppelvlak aangebied word nie — dit kan en hoef dit ook nie te probeer wees nie. Daar moet dus nie soveel opsies op die gebruikerskoppelvlak geplaas word sodat die bladsy beknop of oorvol lyk nie, en ook nie soveel interne klikkruisverwysings dat die gebruiker sy of haar oorspronklike plek of bladsy heeltemal byster raak nie. Net soos wat daar 'n gevaar van 'inligtingdood' bestaan, is daar ook 'n moontlike gevaar van 'klikdood', wat die gebruiker op die ou end laat vra: "Hoe het ek hier gekom? En hoe kom ek weer terug waar ek was?"

Waar die skakels na buitetekste lei — soos die tabel van plaasdiere en hul geluide in figuur 7 — kan die buitetekste as 'n opwipitem ("pop up item") aangebied word wat weer toegemaak kan word sonder dat die gebruiker sy of haar plek verloor en van die bladsy weggeneem word.

Soos reeds in paragraaf 4.2 genoem, speel die leiding wat in die gebruikershandleiding gegee word 'n geweldige belangrike rol in die teikengebruikers se begrip en gebruik van enige buitetekste — dit is noodsaaklik dat hulle weet watter buitetekste tot hul beskikking is, die inhoud wat in daardie buitetekste aangebied word en hoe en wanneer om dié tekste te raadpleeg.

Ander tabelle of lyste wat vir insluiting as buitetekste oorweeg sou kon word, sluit die volgende in:

1. Alfabet (hoof- en kleinletters van die geskrewe alfabet, en vingerspelling met beide 'n video-opname van die hele alfabet van A tot Z, sowel as 'n grafiese voorstelling of lynskets van elke individuele letter)
2. Die Kalender: Dae van die week
3. Die Kalender: Maande van die jaar
4. Seisoene
5. Syfers/Getalle
6. Teenoorgesteldes
7. Kleure
8. Geometriese vorme
9. Klassifiseerders

Die items wat onder nommers 2 tot 7 in die bostaande lys verskyn, vorm tematiese kategorieë binne die klaswoordelys en sal dus reeds op 'n tematiese prentjiesbladsy in die tematiese afdeling van die woordeboek (met die moontlike insluiting van nommer 8, geometriese vorme) behandel word: die gebruiker sal waarskynlik eerder verwag dat hierdie temas, net soos al die ander temas in die woordelyste, eerder 'in' die woordeboek verskyn as 'buite' die woordeboek. Hierdie kategorieë kan egter steeds as buitetekste oorweeg word.

In die geval van die voorgestelde skoolwoordeboek sou maksimumtoegang verseker kon word deur 'n geïntegreerde lys **vingerspelgebare** (hetsy geleksikaliseerde gebare of leenwoorde, indien enige wel in die klaswoordelyste voorkom) in te sluit én al dié gebare in die lys ook as lemmakandidate in die databasis van die woordeboek in te sluit. Die skakel/kruisverwysing na die lys van vingerspelgebare — indien dit ingesluit word — sal dan in die mikrostruktuur van die betrokke leksikale items aangebied word, sodat die gebruiker wat wel so 'n gebaar opsoek, gelei kan word na die volledige tabel van gebare wat op soortgelyke wyse gevorm word. Die lys vingerspelitems kan ingesluit word by die skakel wat in die navigasiekolom na die alfabet lei.

Soortgelyk aan die **eiename** wat in 'n buitetekst in die DASL opgeneem is (sien paragraaf 3), kan 'n lys belangrike en/of algemene eiename wat dikwels voorkom, soos "Nelson Mandela" of "Jacob Zuma", ook as 'n agterteks aangebied word, wat 'n voorbeeld van 'n niegeïntegreerde buitetekst is, aangesien hulle nie die volledige bewerking ontvang wat deur die werklike doel van die woordeboek voorgeskryf word nie.

Net soos in die geval van naamgebare vir persone, word 'n spesifieke gebaar dikwels aan 'n **pleknaam** toegeken, soos aangedui deur die voorbeeld wat hierbo aangegee is. Ook in SASL bestaan daar gebare vir spesifieke plekname — dus 'n gebaar wat, soos naamgebare, nie die hele naam vingerspel nie, maar 'n verkorte vingerspelling kan wees (soos "WK" vir die Wes-Kaap), of 'n selfstandige gebaar soos vir "Mpumalanga" (wat die betekenis van die SiSwatiwoord voorstel, nl. "plek waar die son opkom"). 'n Lys belangrike en/of algemene plekname kan dus ook as 'n niegeïntegreerde buitetekst ingesluit word.

Beide lyste eiename en plekname kom handig te pas in 'n woordeboek waarin die keuse van lemmas op die algemene leksikale items van die gegewe taal fokus, maar waar die leksikograaf items wat dikwels voorkom (of wat probleme met spelling lewer, bv.) wil insluit. Sulke items hoef nie noodwendig dieselfde behandeling te ontvang as wat aan die lemmata in die sentrale lys gegee word nie — 'n agterteks met hierdie items kan slegs 'n beperkte behandeling insluit wat fokus op die inligting wat van toepassing op die behoeftes van die beoogde teikengebruiker is (Gouws en Prinsloo 2005: 60).

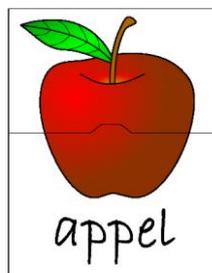
4.6 Speletjies as buitetekste

Bykans elke moderne kind assosieer vandag 'n rekenaar nie (net) met die leerproses nie, maar veral met speletjies wat op die rekenaar gespeel word. Tydens

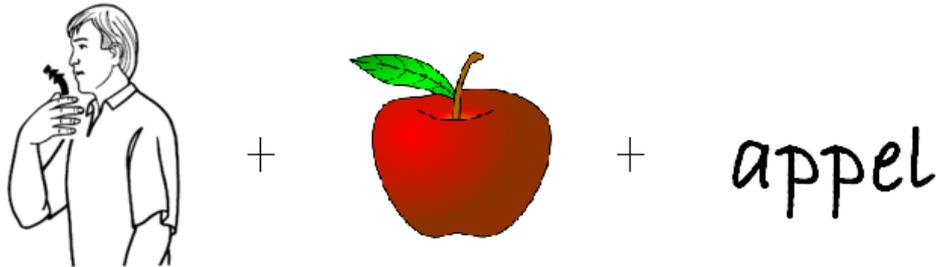
skoolbesoeke is opgemerk hoe selfs voorgraadse Dowe kinders met groot entoesiasme om die rekenaar in die klaskamer saamkoek om speletjies te speel of insette te lewer t.o.v. die 'prestasie' van die kind wat beheer oor die rekenaar muis het. Selfs met net een rekenaar tot die groepie se beskikking het 'n samewerkende speel-/leerproses ontstaan.

Die buitetekste kan ook sekere **aktiwiteite/speletjies** wat om sekere woordgroepe gebou word, huisves: dit is relatief maklik om 'n voltooide JavaScript-programmetjie, byvoorbeeld, van die internet te laai en te gebruik om prentjies (van tematiese kategorieë of items binne die kategorie) in legkaartstukke te verander wat deur die gebruiker bymekaar gepas moet word. So 'n aktiwiteit kan gebruik word om spelvaardighede te oefen — 'n beperkte aantal woorde word opgebreek (vir die jonger, voorskoolse gebruikers in nie meer as twee stukke nie, vir ouer leerders in die grondslagfase waarskynlik meer) en die onderskeie dele moet dan bymekaar gepas word. Die rekenaar gee terugvoer soos "Probeer weer" of "Korrek" (in sowel die geskrewe taal as gebaretaal) afhange van die sukses van die poging.

1. Voorskoolse leerders: sluit prentjie by gebroke legkaartstuk in sodat dit 'n visuele leidraad kan gee (waar geletterdheid nog nie so gevorderd is nie). Die herkenning en passing van die woord kan dus bloot "toevallig" ook wees, maar dit maak nie saak nie, aangesien die aktiwiteit steeds passiewe woordeskat (herkenning i.p.v. produksie) sal bevorder. Produksie is 'n leeruitkoms in die voorskoolse fase slegs vir die oorskryf van 'n enkele woord op 'n slag, nie teksproduksie in die gewone sin van die woord (dus een wat die produksie van volsinne behels nie). Die klem val in die voorskoolse klas meer op die ken van die gebaar (dus produksie en resepsie) en die herkenning van die woord (resepsie).
2. Grondslagfaseleerders: Plaas die regte gebaar (op 'n lynskets of selfs op 'n video) by die passende prentjie en/of woord. Die hele geskrewe woord moet dus nou in isolasie by die korrekte gebaar of prentjie gepas word. Die aktiwiteit fokus steeds op die passiewe woordeskat, maar daar is nie meer dieselfde mate van visuele leiding wat verskaf word nie.

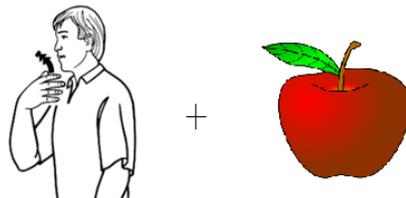


Figuur 12: Eenvoudige legkaartspeletjie waar twee dele van die legkaart bymekaar gepas moet word om die prentjie te voltooi



Figuur 13: Meer gevorderde tipe legkaartspeletjie waar die gebaar by die korrekte prentjie en/of woord gepas moet word (<http://nzsl.vuw.ac.nz/signs/4199> — die lynskets is ter illustrasie vanaf die NZSL-webtuiste geneem en is nie noodwendig dieselfde as die SASL-gebaar vir "appel" nie)

Pleks van 'n legkaart, kan die verskillende komponente wat saam hoort (gebaar plus prentjie plus geskrewe woord) in 'n tabel aangebring word (d.m.v. "drag and drop") om dan terugvoer van die rekenaar te ontvang, m.a.w. "Probeer weer" of "Korrek" (weereens in albei tale). Indien 'n leerder byvoorbeeld slegs een of twee van die komponente (soos gebaar en prentjie) korrek bymekaar kan pas en nie seker is wat die geskrewe vorm is nie, kan 'n "Help"-knoppie die gebruiker in die regte rigting aanpor deur óf onmiddellik die hele woord in te vul óf die letters een na die ander in te vul totdat die gebruiker (hopelik) die hele woord self onthou en verder kan invul.



Figuur 14: Gebaar plus prentjie



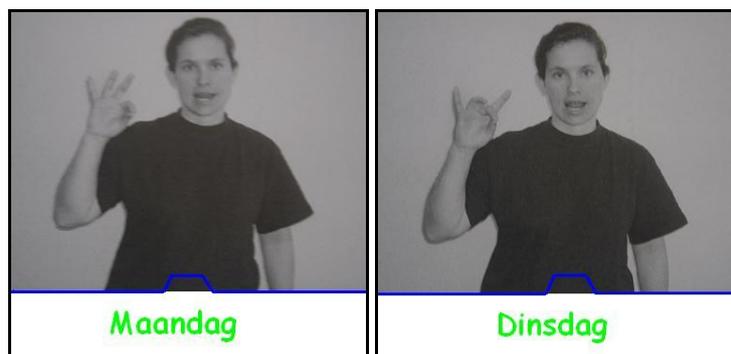
Figuur 15: Gebaar plus geskrewe woord

Ander aktiwiteite/oefeninge wat op 'n soortgelyke manier gedoen kan word, is:

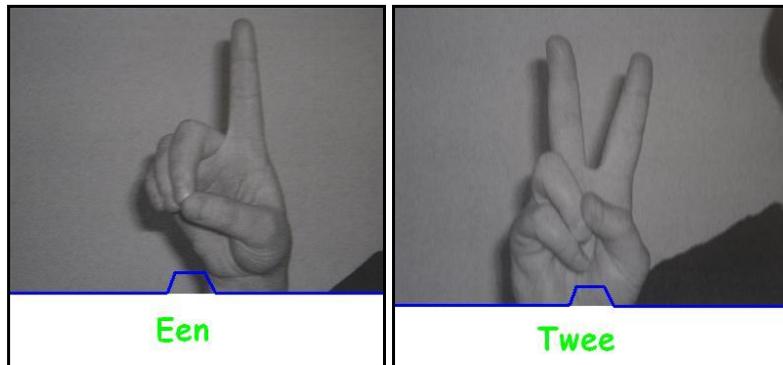
- Teenoorgesteldes
- Vorms
- Kleure
- Syfers
- Dae en Maande
- Voorsetsels
- Soek woorde in 'n ander taal (bv. tweede geskrewe taal) (die pas van die geskrewe woord by 'n video of skets van die gebaar is eintlik reeds 'n "soek die woord in 'n ander taal"-aktiwiteit).



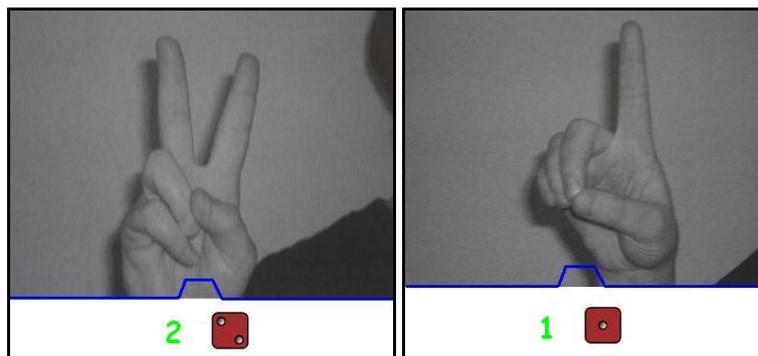
Figuur 16: Voorbeelde van teenoorgesteldes wat bymekaar gepas word



Figuur 17: Voorbeelde van die gebaar vir 'n dag van die week wat by die geskrewe vorm gepas word



Figuur 18: Voorbeelde van gebare vir getalle wat by die geskrewe vorm gepas word



Figuur 19: Voorbeelde van gebare vir getalle wat by die geskrewe syfer tesame met 'n grafiese voorstelling van die getal se waarde (t.w.v. konseptualisering) gepas word

Hierdie speletjies kan dan in 'n gesamentlike groep as buitetekst of 'kategorie' verskaf word deur 'n skakel na die lys speletjies te verskaf op dieselfde bladsy wat die skakels na al die tematiese kategorieë verskaf — sien die skematiese uitleg hiervan in Fourie 2013: 315-325. Skakels kan ook (hierdie of enige ander) buitetekste verbind met relevante inskrywings in die mikrostruktuur van artikels wat via gebaar of woord gesoek word. Vir 'n bespreking van die verskillende soekmetodes in die voorgestelde woordeboek, sien Fourie 2013: 303-338.

4.7 Afwisselende gebruik van verdere moontlike buitetekste

Die elektroniese woordeboek wat aanlyn aangebied word, bied ook die unieke moontlikheid om 'n dinamiese omgewing t.o.v. die buitetekste te skep — hierdie tekste kan gereeld gewissel en aangepas word na gelang van die behoeftes

wat die gebruikers ook al op 'n spesifieke oomblik het en die leksikograaf het die vryheid om aan die hand van gebruikersbehoefte en –terugvoer, asook eie goeddunke, die verskeidenheid buitetekste wat aangebied word, te meng en te pas. Alle moontlike interne en eksterne buitetekste hoef dus nie op een stadium aangebied te word nie, aangesien dit waarskynlik net tot informasie- en selfs skakel-/klikdood sal lei. Die buitetekste wat aangebied word en die aktiwiteit wat om dit gebruik word, moet ouderdomspesifiek en ook gepas vir die gebruiker se vaardigheidsvlak wees. 'n Kieslys moet aan gebruikers leiding gee oor watter buitetekste in die databasis aan hulle beskikbaar is.

Skakels na buitetekste/bladsye oor die **alfabet, getalle en klassifiseerders** kan as die minimum vir buitetekste (benewens die verpligte instruksies aan die gebruiker, en die minigrammatikas van die twee tale) beskou word en dit word dus aanbeveel dat hierdie tekste in die voorgestelde woordeboek ingesluit word. Ander buitetekste soos hierbo bespreek, bv. die tabelle van **dieregeluide** en **dierenname**, asook die **trappe van vergelyking, eiename en plekname, idioome** en **speletjies** kan ook vir insluiting oorweeg word. Verdere skakels na inligting soos 'n lys **mates en gewigte**, die land se **provinsies en hoofstede**, en 'n lys van die name van die **tale** wat in die land gebruik word, sal ook tot die kognitiewe funksies van die buitetekste bydra en gebruikers toelaat om die woordeboek ook as skakel na ander ensiklopediese inligting te benut sodat die omvang van die woordeboek se moontlikheid as naslaanbron eintlik onbeperk is.

'n Skakel na die **kontakbesonderhede** van die woordeboek se samestellers sal dit vir gebruikers moontlik maak om terugvoer oor hul ervaring van en behoeftes ten opsigte van die woordeboek te lewer, wat die leksikograaf verder in staat sal stel om 'n produk te lewer wat vir die gebruikers pasgemaak is. Dit is veral belangrik dat ook die onderwysers moet kan terugvoer gee, aangesien die woordeboek hoofsaaklik daarop gemik is om in klassituasies gebruik te word en as aanvulling (indien nie selfs as plaasvervanging) vir die geskrewe klaswoordelyste te dien.

'n Lysie skakels na ander **woordeboeke**, in gebaretaal of die geskrewe taal, kan ook ingesluit word.

Die leksikografiese span kan ook weekliks of selfs daaglik 'n **gebaar-van-die-week/dag** kies op grond van kriteria soos belangrike en/of aktuele gebeure plaaslik of elders. Soos wat die gebruikers mettertyd besef dat die inhoud van die woordeboek dinamies is en gereeld vir nuwe 'verrassings' sorg, behoort hul belangstelling en nuuskierigheid ook geprikkel te word om gereeld na die woordeboek terug te keer. Deur so 'n proses kan die geleidelike ontwikkeling van 'n woordeboekkultuur ook aangewakker word.

Nuwe soorte buitetekste wat oorweeg kan word, sluit die volgende in:

1. 'n skakel na 'n **aanlyn forum** waar gebruikers gebare na 'n bediener kan oplaai vir moderering en moontlike uiteindelijke insluiting in die woordeboek;
2. 'n tipe "**Maak jou eie woordelys**"-funksie waarvoor daar by elke inskry-

wing 'n skakel verskaf word. 'n Manipuleerbare lys gebare (met die geskrewe woord en 'n lynskets van die gebaar) word dan saamgestel en kan op papier uitgedruk word of selfs elektronies gestoor word as "My woordeboek" sodat die gebruiker met elke sessie kan terugkeer na die gebare wat hy of sy om watter redes ook al uit die woordeboek gekies en opgesit het. Met "manipuleerbaar" word bedoel dat die gebruiker die volgorde van die gebare op die woordelys self kan aanpas deur die gebare-skets te heen en weer oor die skerm te sleep met die muis ("drag and drop"). Die gebruiker het d.m.v. so 'n funksie die moontlikheid om sy of haar eie idiolekwoordeboek te skep;

3. skakels na die skool se **webblad**, NID (Nasionale Instituut vir Doves, wat ook in Worcester geleë is) se webblad en Facebook-bladsy, DEAFSA, Dowe sportverenigings en sosiale klubs, ens., wat by die volgende afdeling oor kultuur aansluit;
4. 'n Afdeling oor die **Dowe gemeenskap** of **kultuur**, wat video's van stories en gedigte, ens. in gebaretaal aanbied, asook gedragskodes binne die kultuur t.o.v. dinge soos die etiket wat die maniere bepaal waarop iemand se aandag gekry kan word, beurtneemings in 'n gesprek en onderbreking, hoe en wanneer mens 'deur' 'n gesprek mag loop, die vereistes t.o.v. beligting in 'n vertrek waar Doves met mekaar of met horendes kommunikeer, etiket en moontlike strategieë wanneer horendes wat nie gebaretaal ken nie met Doves kommunikeer d.m.v. liples, geskrewe notas, of dergelike, hoe applous in gebaretaal lyk, ens. Deur sulke inligting in 'n buiteteks beskikbaar te maak, word die Dowe gemeenskap ook kultureel gekontekstualiseer, en help dit die horende gebruiker van die woordeboek om ook iets van die nuwe/ander kultuur te leer. Dit is ook iets wat binne die kognitiewe funksie van die woordeboek val (Gouws 2009b). 'n Lys gebare wat spesifiek aan Dowe kultuur is, kan ook aangebied word, met skakels vanuit die buiteteks na die betrokke artikels;
5. die woordeboek kan selfs 'n skakel na sy **eie Facebook-bladsy** hê waarop o.a. die gebare van die week/dag (sien hierbo) geplaas kan word. Deur middel van gewilde sosiale media soos Facebook kan die woordeboek moontlik 'n al groter groep gebruikers bereik en bewustheid van die woordeboek help versprei.

Die idee is hoegenaamd nie dat **al** die bogenoemde buitetekste op een slag aangewend word nie, veral nie in 'n grondslagfasewoordeboek nie — die bespreking dien slegs as 'n aanduiding van die opsies wat oorweeg kan word. Dit word weereens benadruk dat buitetekste gekies moet word op grond van hul nut en gepastheid vir 'n spesifieke gebruikersgroep, en dat sommige aanvullende tekste op 'n tipe rotasiebasis aangewend kan word. Al is die moontlikhede legio mag die gebruiker nooit deur die aanbod oorweldig word nie.

5. Slot

In hierdie artikel oor minigrammatikas en ander buitetekste in die raamstruktuur van die beplande woordeboek was die fokus die inhoud van minigrammatikas van beide SASL en Afrikaans as bewerkingstale en bewerkte tale en die hoofstuk het 'n oorsig gegee van onderwerpe wat in die minigrammatikas hanteer behoort te word of kan word. Ten spyte daarvan dat die artikel voorstelle maak oor die tipe grammatikale inligting wat t.o.v. SASL ingesluit behoort te word, is verdere navorsing oor die besonderhede van die grammatika SASL steeds nodig, aangesien 'n volledige grammatikale beskrywing van die taal steeds ontbreek. Voorbeelde uit ander gebaretale is daarom dikwels nodig om 'n punt te illustreer.

'n Reeks voorbeelde van die gebruik van verskillende soorte buitetekste in beide gedrukte woordeboeke en elektroniese gebaretaalboeke is ook ondersoek en 'n wye verskeidenheid voorstelle is gemaak vir die tweetalige elektroniese woordeboek van SASL en Afrikaans vir die grondslagfase wat uit Fourie (2013) se konsepmodel sou kon ontstaan.

Eindnote

1. Daar is besluit om deurgaans by die (Engelse) algemeen bekende afkortings vir verskillende Gebaretale te hou: so word SASL (South African Sign Language) dus gebruik eerder as SAGT (Suid-Afrikaanse Gebaretaal); ASL (American Sign Language) word gebruik eerder as AGT (Amerikaanse Gebaretaal). Hierdie keuse is gemaak omdat die bestaande literatuur oor gebaretale gewoonlik in Engels is en die meeste lesers reeds met die Engelse afkortings vertrou is.
2. Die gebruik van die kleinletter in doof dui op oudiologiese of patologiese doofheid, terwyl die hoofletter in Doof dui op kulturele Doofheid.
3. Die kleinste kontrasterende eenhede van gebaretaalanalise, soortgelyk aan foneme in gesproke tale.

Bronnelys

- Al-Ajmi, H.** 2001. The Role of the Introductory Matter in Bilingual Dictionaries of English and Arabic. *Lexikos* 11: 60-70.
- Bergenholtz, H. en R.H. Gouws.** 2010. A New Perspective on the Access Process. *Hermes. Journal of Language and Communication Studies* 44: 103-127.
- Bergenholtz, H., S. Nielsen en S. Tarp (Reds.).** 2009. *Lexicography at a Crossroads: Dictionaries and Encyclopaedias Today, Lexicographical Tools Tomorrow*. Linguistic Insights. Studies in Language and Communication 90. Bern: Peter Lang.
- Bergenholtz, H., S. Tarp en H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Reds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Vol. 2: 1762-1832*. Berlyn: De Gruyter.

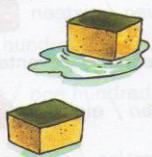
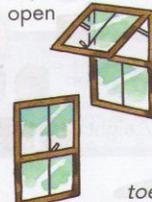
- Brennan, M.** 1992. The Visual World of British Sign Language: An Introduction. Brien, D. (Red.). 1992. *Dictionary of British Sign Language/English*: 1-133. Londen: Faber&Faber.
- Brien, D. (Red.).** 1992. *Dictionary of British Sign Language/English*. Londen: Faber&Faber.
- Fourie, H.** 2013. 'n Leksikografiese model vir 'n elektroniese tweetalige grondslagfasewoordeboek van Suid-Afrikaanse Gebaretaal en Afrikaans. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- Gallardo, A.** 1980. Dictionaries and the Standardization Process. Zgusta, L. (Red.). 1980. *Theory and Method in Lexicography*: 59-69. Columbia: Hornbeam Press.
- Gouws, R.H.** 2001. Lexicographic Training: Approaches and Topics/Formation lexicographique: approches et thèmes.. Emejelu, J.D. (Red.). 2001. *Éléments de lexicographie gabonaise. Tome I*: 58-94. New York: Jimacs-Hillman.
- Gouws, R.H.** 2009a. The Integrated Outer Texts in Recent English and German Learners' Dictionaries: A Critical Comparison. *Lexicographica* 25: 67-90. Tübingen: Max Niemeyer Verlag.
- Gouws, R.H.** 2009b. Geïntegreerde woordeboekgebruik van vakwoordeboeke vir aanleerders. *Lexikos* 19: 72-93.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PreSS.
- Gouws, R., L. Potgieter en S. Burgess (Reds.).** 2010. *Longman Grondslagfasewoordeboek Afrikaans/Engels*. Kaapstad: Maskew Miller Longman.
- Hausmann, F.J. en H.E. Wiegand.** 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie*: 328-360. Berlyn: Walter de Gruyter.
- Hiles, L. en I. Redelinghuijs (Reds.).** 2010³. *Oxford First Bilingual Dictionary: Afrikaans + English*. Kaapstad: Oxford University Press.
- Howard, S.** 2008. *Finger Talk: South African Sign Language (SASL) Dictionary*. Gillitts: Fulton School for the Deaf.
- Mavoungou, P.A.** 2002. *Metalexical Criteria for the Compilation of a Trilingual Dictionary: Yilumbu-English-French*. Ongepubliseerde D.Litt.-proefskrif. Stellenbosch: Universiteit Stellenbosch.
- McKee D., R. McKee, S. Pivac Alexander, M. Vale en L. Pivac (Reds.).** 2011. *Online Dictionary of New Zealand Sign Language*. Deaf Studies Research Unit, Victoria University of Wellington, New Zealand. <http://nzsl.vuw.ac.nz/nzsl>.
- Penn, C., D. Ogilvy-Foreman, D. Simmons and M. Anderson-Forbes.** 1992. *Dictionary of Southern African Signs for Communicating with the Deaf*. Deel 1. Pretoria: Raad vir Geesteswetenskaplike Navorsing.
- Procter, P. (Ed.).** 1978. *Longman Dictionary of Contemporary English*. Harlow: Longman.
- Stokoe, W.C., D.C. Casterline en C.G. Croneberg.** 1976. *A Dictionary of American Sign Language on Linguistic Principles*. Nuwe uitgawe. Washington DC: Gallaudet University Press.
- Valli, C. (Red.).** 2005. *The Gallaudet Dictionary of American Sign Language*. Washington, DC: Gallaudet University Press.

Addendum A: Voorbeeld van buitetekste in die LGFW (Gouws et al. 2010: 1)



Addendum B: Voorbeeld van buitetekst oor teenoorgesteldes in die OFBD (Hiles en Redelinghuijs 2010: 39)

Teenoorgesteldes / Opposites

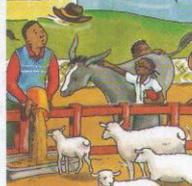
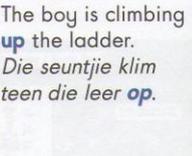
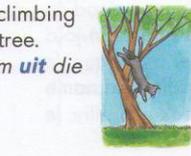
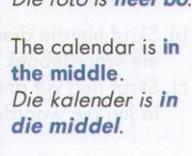
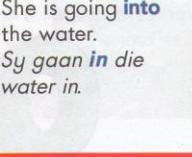
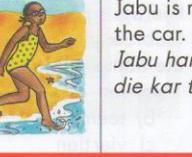
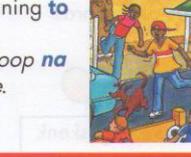
<p>donker dark</p>  <p>lig light</p>	<p>eenders same</p>  <p>verskillend different</p>	<p>gelukkig happy</p>  <p>treurig sad</p>	<p>groot big</p>  <p>klein small</p>	<p>hoog high</p>  <p>laag low</p>
<p>kort short</p>  <p>lank tall</p>	<p>nat wet</p>  <p>droog dry</p>	<p>nuut new</p>  <p>oud old</p>	<p>oop open</p>  <p>toe closed</p>	<p>skoon clean</p>  <p>vuil dirty</p>
<p>swaar heavy</p>  <p>lig light</p>	<p>vet fat</p>  <p>maer thin</p>	<p>vinnig fast</p>  <p>stadig slow</p>	<p>vol full</p>  <p>leeg empty</p>	<p>warm hot</p>  <p>koud cold</p>

*My hare is droog, maar joune is nog nat.
Jou glas is vol, maar myne is leeg.*

*My hair is dry, but yours is still wet.
Your glass is full, but mine is empty.*

Addendum C: Voorbeeld van buitetekste oor voorsetsels in die OFBD (Hiles en Redelinghuys 2010: 47)

 **Waar is dit? / Where is it?** 

 <p>The bird is on the roof. <i>Die voël is op die dak.</i></p>	 <p>There is cool drink in the glasses. <i>Daar is koeldrank in die glase.</i></p>	 <p>The T-shirt is under the bed. <i>Die T-hemp is onder die bed.</i></p>	 <p>The light is above the television. <i>Die lig is bo die televisie.</i></p>
 <p>The nurse is in front of the clinic. <i>Die verpleegster is voor die kliniek.</i></p>	 <p>The donkey is next to the farmer. <i>Die donkie is langs die boer.</i></p>	 <p>Who is behind the door? <i>Wie is agter die deur?</i></p>	 <p>The photograph is at the top. <i>Die foto is heel bo.</i></p>
Waar gaan hulle heen? / Where are they going?			
 <p>The boy is climbing up the ladder. <i>Die seuntjie klim teen die leer op.</i></p>	 <p>The cat is climbing down the tree. <i>Die kat klim uit die boom.</i></p>	 <p>The calendar is in the middle. <i>Die kalender is in die middel.</i></p>	 <p>The list is at the bottom. <i>Die lys is heel onder.</i></p>
 <p>She is going into the water. <i>Sy gaan in die water in.</i></p>	 <p>Jabu is running to the car. <i>Jabu hardloop na die kar toe.</i></p>	 <p></p>	<p></p>

Addendum D: Gebruikersgids tot die uitleg van die BSL-afdeling van die DBSL (Brien 1992: xiv)

How to use the Dictionary
[BSL section layout]

807 $\emptyset \checkmark < \Lambda$ $\overset{\circ}{\Lambda}$

808 $\Gamma \checkmark \tau \Lambda$ $\overset{\circ}{\Lambda}$

809 $\Gamma \checkmark \tau \Lambda$ $\overset{\circ}{\Lambda}$

1 Handshape illustration

The handshape illustration on the left-hand page shows the handshape used as the dominant (i.e. right) hand for the signs on these pages; see page xvii–xix for more details. Some signs show minor variations; these are discussed in the introduction to each of the handshape sections.

2 Entry number

Each sign entry is given a number for reference. This number helps you to find more information about this sign on the facing page and also in the *English (Guide to meaning)* section in the second half of this volume: see page xvi for further information.

3 Notation

Sign entries are notated using the BSL notation system developed by linguists in the UK. This is a scientific recording system which gives detailed information on paper about the production of the signs of a visual-gestural language (see pages xvii–xvii for more details about the system of artwork). It is used here to work alongside the *Photographs* and *Production guide* in giving a complete explanation of how to produce each sign.

4 Photographs

Sign entries consist of one or more images, which must be 'read' from left to right or top to bottom. These have artwork added to show the movements made as the sign is produced (see pages xxviii–xxix for further information). To get maximum benefit from the range of information presented in the *BSL* section, we recommended that readers al-

ways use both images and text (plus *Notation*, for those who are familiar with that system) together to find out how the sign is produced.

5 Handshape symbol

Each handshape has its own symbol in the notation system (see pages xvii–xix). The symbol shown here relates exactly to the *Handshape illustration*. Note especially that neither the symbols nor the handshapes to which they relate are, in fact, representations of letters of the English alphabet, i.e. *A* is not related to 'a'.

6 Handshape description

To help you make the handshapes accurately, we give a short description of the typical appearance of the dominant handshape (i.e. the right hand) in all the *Photographs* on this page. This description matches up with the *Notation* symbol beside it, and the *Handshape illustration* on the opposite page.

7 Column headings

The information text is displayed in four columns. The first three columns tell you step-by-step how to produce the manual elements of the sign. The fourth column offers a number of English *Glosses* that correspond, in various contexts, to the sign entry.

8 Notes

Text entries may include additional notes (see page xxxi), which are of two kinds. The symbols *, † and ‡ are used when the information that follows is supplementary—e.g. if a sign is a 'Regional sign', as in entry

Addendum E: Gebruikersgids tot die uitleg van die BSL-afdeling van die DBSL (vervolg) (Brien 1992: xv)

How to use the Dictionary
[BSL section layout]

	5	6	7	8	9	10	11	12	13	14	15	
	The best index and middle fingers are extended from the fist and spread apart.										807-809	
<i>Handshape and orientation</i>	<i>Location, hand arrangement and contact</i>				<i>Movement</i>				<i>Glosses</i>			
<p>The V hand is held with the palm facing left (if the hand were opened, the fingers would point up).</p> <p><i>Notes</i></p> <p>Regional sign.</p> <p>1 If the movement is changed to a short repeated downward movement, the sign produced is metro, subway, tube, underground, underground train.</p>	<p>The hand is held in front of the shoulder.</p>				<p>The hand makes a short movement away from the signer.</p>				<p>bus* metro *1 subway *1 tube *1 underground *1 underground train *1</p>			
<p>The V hand is held with the palm facing towards the signer (if the hand were opened, the fingers would point up).</p> <p><i>Non-manual features</i></p> <p>The tips are rounded and pushed forward.</p>	<p>The hand is held with the extended fingers in front of the eyes.</p>				<p>The hand makes two short movements from side to side.</p>				<p>blind as a bat miss</p>			
<p>The V hand is held with the palm facing towards the signer (if the hand were opened, the fingers would point up).</p>	<p>The hand is held with the extended fingers in front of the eyes.</p>				<p>The hand makes short twisting movements from side to side.</p>				<p>blind blind person</p>			

807 above – or explains how a sign may be modified. The symbols ¹, ², ³ and ⁴ are used to explain that a sign can only be produced following the instructions in the *Notes*.

9 Non-manual features
If non-manual features are specified, remember that these are integral parts of the sign. The non-manual features are produced at the same time as the manual aspects of the sign. As it has not always been possible to capture the features in the *Photographs*, you should always check here for a description of the non-manual features. See pages xxiv–xxvii for more details of these features.

10 Production guide one
The information in the first column tells you (using *Notation* symbols) which handshapes are used at the start of the sign and how they are orientated in space. Note that the production specification is always given from the signer's own point of view. See page xxx for further information.

11 Production guide two
The second column of texts explains where the hands are held and how they are positioned relative to each other. Follow the information in this and the first column and your hands will be in position to begin the sign (see page xxx).

12 Production guide three
In the third column, you will find a description of the movement made

by the hands in the production of the sign, including any changes in the shape of the hands (see page xxx).

13 Glosses
The glosses column contains a number of entries that can be thought of as possible parallels to the sign given opposite. Two types of additional information are indicated in this column. Firstly, we have used the symbols *, † and ‡ to give supplementary information or to show that the sign with this gloss may be modified in certain ways as indicated in the *Notes*. Secondly, we have used the symbols ¹, ², ³ and ⁴ when the sign with this gloss can only be produced following the instructions in the *Notes* (see page xxxi for details).

14 Entry number repeated
Numbers (corresponding to the *Entry numbers* on the facing page) are located at the side of the text. Each number provides a direct link between a section of descriptive text here, the *Photographs* (plus *Notation*) opposite and an entry or entries in the *English (Guide to meaning)* section. See pages xxxii–xxxiv for further information.

15 Page entry contents
In the top corner of the page, you will find numbers showing the range of entries contained on these two pages, to help you locate quickly the information you need (see pages xxxii–xxxiv for more details).

[xv]

Addendum F: Gebruikersgids tot die uitleg van die Engelse gids tot betekenis-afdeling van die DBSL (Brien 1992: xvi)

*How to use the Dictionary
(English (Guide to meaning) section layout)*

The diagram illustrates the layout of a dictionary page. At the top, a horizontal line contains the words 'address' on the left and 'advocate' on the right. Eight numbered arrows point from this line to specific features in the dictionary entries below:

- 1** points to the word 'address' in the left margin.
- 2** points to the first entry for 'address'.
- 3** points to the second entry for 'address'.
- 4** points to the reference number '23, 964'.
- 5** points to the word 'admit'.
- 6** points to the first sense of 'admit'.
- 7** points to the second sense of 'admit'.
- 8** points to the word 'advocate' in the right margin.

address
An address is a formal signed or spoken presentation that someone makes to a group of people, often on a special occasion. *e.g. She gave the opening address at the isla symposium.*

address
Your address is the number of the house, the name of the street, and the town where you live or work. *e.g. Write down your name and ad*

admit
1247
1 If someone or something is **admitted** to a place, they are allowed to enter it. *e.g. Junior members of staff can be admitted.*
2 If someone is **admitted** to hospital, they are kept there until they are well enough to go home. *e.g. He was admitted to hospital with an ulcerated leg.*
3 If someone is **admitted** to an organisation or group, they are allowed to join it. *e.g. He was*

advertisement
1205
An advertisement (or advert for short) is an announcement in a newspaper, on television, or on a poster about a product, event, job vacancy, etc. *e.g. We've received several replies to our advertisement.*

advocacy
229
Advocacy of a particular action or plan is the act of supporting it publicly. *e.g. The BDA has*

1 First word
In the top left-hand corner of every page in the *English (Guide to meaning)* section of this Dictionary, you will find a word printed in black letters like this. It is the first item that has an entry on this page. When you want to look through the book quickly, you can easily find the page you want by keeping an eye on this corner of the page (see also *Last word*, below).

2 Headword
The glosses listed in the column on the right-hand side of each right-hand page in the *BSL* section are used in this section of the Dictionary as headwords. The headword (printed in bold letters, like this) introduces the explanation of the meaning. The headword (and all information laid out in the text below it) actually relates to each one of the signs whose *Reference numbers* appear beside it. If a headword has two or more distinct meanings, each relating to a different sign or signs, then the headword is repeated as a new entry and the relevant sign *Reference numbers* are listed beside it. For example, **address** appears twice as the headword on this page. The first entry is linked to signs 1335 and 872 and its meaning relates to 'a formal presentation'. The second entry is linked to sign 1886 and its meaning relates to 'the place where you live or work'.

3 Reference number
The reference numbers used here direct the reader to the appropriate corresponding entry or entries in the *BSL* section of the book. For instance, if the number 1335 appears beside the explanation, then sign entry 1335 will show a *BSL* sign (photograph, notation, description and relevant notes) whose meaning relates to this explanation. Likewise, if you know the entry number and gloss of a particular *BSL* sign and you turn from the *BSL* section to the *English (Guide to meaning)* section, remember that only those explanations linked to the relevant reference number will match the *BSL* sign you have in mind.

4 Senses and uses
If there are two or more meanings associated with the sign or signs (see *Reference number*) listed in relation to the given headword, then these are entered as separate numbered paragraphs.

5 Headword repeated
In the explanation of the meaning, the *Headword* (or a form of the *Headword* adapted to be as clear as possible in the context of the explanation) is printed in bold letters, like this.

6 Explanation of meaning
The *Headword* is always explained in full sentences. This means that you can see how it is used in context. Remember, the explanation deals primarily with the meaning of the *BSL* sign or signs identified by the *Reference numbers* listed beside the explanation, and only secondarily with the meaning of the English word used as the *Headword*. The signs are listed here according to their glosses so that they can be arranged under *Headwords* in alphabetical order (easier to look up) and especially so that the correspondence between the *BSL* signs and the English words—not necessarily a one-to-one relationship—can be made clear. One *BSL* sign can often be translated by any of several different words of English; one English word can often be translated by any of several different signs of *BSL*.
If we have chosen accurate English glosses—that is, words and phrases which can comfortably be used when translating the sign into English—to use as *Headwords*, then it is logical that the explanation of the meaning of the *BSL* sign will also be appropriate for this sense of the English word. Even if the English word has many other meanings, these will not be recorded in this Dictionary unless corresponding *BSL* signs are listed in the *BSL* section of the book: that is why this is only secondarily a guide to the meaning of the English words.

7 Example sentence
The example sentences in this Dictionary appear in italics (like this) after each *Explanation of meaning*. The examples have been chosen to show typical patterns and structures and, wherever possible, are taken from real language contexts. Although the examples are presented in English, they have been chosen with great care to be applicable both to the English word and to the *BSL* sign or signs on which the explanation focuses (i.e. the sign or signs whose *Reference numbers* appear above the *Explanation of meaning*).
Our aim is that the example sentence should represent both a structure of English and a potential sentence of *BSL*, if the written sentence were translated accurately. We claim that a *BSL* translation of the sentence could use the particular sign referred to, and could be idiomatic and grammatical if this were done.

8 Last word
In the top right-hand corner of every page in the *English (Guide to meaning)* section of this Dictionary, you will find a word printed in black letters like this. It is the last item that has an entry on this page (see also *First word*, above).

[xvi]

Article Structures: Moving from Printed to e-Dictionaries*

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

Abstract: By means of an overview of certain aspects of article structures in printed dictionaries and with reference to some examples from e-dictionaries a number of features of article structures in e-dictionaries are discussed. Reference is made to the positioning of articles in article stretches and functional partial article stretches. Different structural components of articles, i.e. text segments, comments and search zones are distinguished. The increased role of data-identifying entries as a type of non-typographical structural indicator in e-dictionaries receives attention as well as the fact that the traditional division of an article in two comments, typically a comment on form and a comment on semantics, cannot merely be maintained. The value of the cohesion resulting from the use of comments in printed dictionaries is much more restricted in e-dictionaries. The use of search zones and rapid access to these zones have a much more important role in the article structure of e-dictionaries. In the planning of e-dictionaries provision needs to be made for a multi-layered article structure with screenshots that display the data in a variety of search zones. Access to these search zones goes via structural indicators in an opening or further screenshot. Provision needs to be made for one lemma to occur in a comprehensive article but also in a number of restricted articles that can be retrieved from the same database. Users should also have the opportunity to design their own user profile that will allow them to consult dictionary articles structured according to their specific needs.

Keywords: ARTICLE STRUCTURE, COMMENT, COMPREHENSIVE ARTICLE, DATA-IDENTIFYING ENTRY, e-DICTIONARIES, ITEM, MULTI-LAYERED ARTICLES, NON-TYPOGRAPHICAL STRUCTURAL INDICATOR, PARTIAL ARTICLE STRETCH, RESTRICTED ARTICLES, SCREENSHOT, SEARCH ZONE, TEXT SEGMENTS, USER PROFILE

Opsomming: Artikelstrukture: vanaf gedrukte na e-woordeboeke. Aan die hand van 'n oorsig oor bepaalde aspekte van artikelstrukture in gedrukte woordeboeke en met verwysing na 'n aantal voorbeelde uit e-woordeboeke word sekere kenmerke van artikelstrukture in e-woordeboeke bespreek. Daar word gekyk na die plasing van artikels in artikeltrajekte asook in funksionele artikeldeeltrajekte. Verskillende struktuurkomponente van artikels, te wete tekssegmente, kommentare en soeksones word onderskei. Die groter rol van data-identifiserende inskrywings as tipe nietipografiese struktuurmerker in e-woordeboeke kry aandag asook die feit dat die tradisionele verdeling in twee kommentare, tipies 'n vormkommentaar en 'n semantiese kommentaar, nie

* This article was the basis of a shortened version presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

sonder meer gehandhaaf word nie. Die waarde van die samehang wat kommentare in gedrukte woordeboeke meebring, is veel beperkter in e-woordeboeke. Die gebruik van soeksones en die kits-toegang daartoe speel 'n veel belangriker rol in die artikelstruktuur van e-woordeboeke. Vir e-woordeboeke moet 'n veelvlakkige artikelstruktuur met skermskote wat die data in 'n verskeidenheid soeksones vertoon, beplan word. Toegang tot soeksones geskied via struktuurmerkers in 'n openings- of verdere skermskoot. Daar moet daarvoor voorsiening gemaak word dat een lemma in 'n omvattende artikel maar ook in 'n verskeidenheid beperkte artikels kan optree wat aan dieselfde databasis onttrek word. Gebruikers moet ook die kans kry om met behulp van die opstel van 'n eie gebruikersprofiel artikels te kan raadpleeg wat in terme van hulle eie behoeftes gestruktureer is.

Sleutelwoorde: AANDUIDER, ARTIKELDEELTRAJEK, ARTIKELSTRUKTUUR, BEPERKTE ARTIKELS, DATA-IDENTIFISERENDE INSKRYWING, e-WOORDEBOEKE, GEBRUIKERSPROFIEL, KOMMENTAAR, NIETIPOGRAFISE STRUKTUURMERKER, OMVATTENDE ARTIKELS, SKERMSKOOT, SOEKSONE, TEKSSEGMENTE, VEELVLAKKIGE ARTIKELS

1. Introduction

Within the field of theoretical lexicography research of the past decades has focused on various topics, including the contents of dictionaries, dictionary structures and lexicographic functions. Theoretical lexicography only emerged long after the lexicographic practice had been firmly established. In many of the developments in theoretical lexicography metalexicographers too often only looked backwards and in their theoretical approaches discussed the contents and structures of dictionaries of the past. Too seldom did they utilise their position to develop models for dictionaries of the future. This assignment was left in the hands of practical lexicographers. To ensure successful lexicographic endeavours, both in terms of theoretical developments and practical dictionaries, it is of paramount importance to maintain a strong and interactive relation between theoretical and practical lexicographers. Neither theoretical nor practical lexicographers should work isolated from the other party. The most important assignment of metalexicographers is not merely to devise models they deem appropriate for the lexicographic practice but rather to consult with practical lexicographers and only then to make suggestions that could enhance the lexicographic practice. As experts working in the field of an independent discipline metalexicographers should use their contact with the lexicographic practice to be innovative and to do research that can result in bringing new dictionary models and designs to the fore that can lead the way to new lexicographic dispensations.

The emergence of e-lexicography has presented metalexicographers with such an opportunity. Prevailing models, proposals and discussions in lexicographic theory were primarily directed at printed dictionaries. The golden opportunity to lead the way in devising models for e-dictionaries was unfortunately not utilised sufficiently by metalexicographers. Yet again too many lexicographic e-products were developed without any influence from innovative

theoretical suggestions. Where practical lexicographers did utilise lexicographic theory it often was a version that had been devised for printed dictionaries. As a result current e-dictionaries too often do not live up to the expectations of their users. A better theoretical basis is needed and that should result in better e-dictionaries.

It is necessary to devise theoretical models for e-dictionaries that focus on critical areas like the data to be included in these dictionaries, the structures to present and accommodate the data, the functions of these dictionaries and the way they should respond to the needs of their target users. Dictionary structures, one of the prominent features of a general theory of lexicography, need to be re-investigated in order to determine which structures are needed in e-dictionaries. The importance of dictionary structures for the successful distribution and presentation of data should never be underestimated. This does not only apply to printed dictionaries but also to e-dictionaries. One of the noticeable features of too many e-dictionaries is a lack of well-planned structures. As a result these dictionaries do not allow their users an optimal access to the data or a comprehensive retrieval of information from the data on offer. Metalexicographers need to pay much more attention to an adaptation of lexicographic theory for application to e-dictionaries. One such aspect that needs attention is the formulation and application of dictionary structures in e-dictionaries. Some structures found in printed dictionaries should be retained, some should be changed, and some should not be used in e-dictionaries, whereas some new structures may have to be introduced.

This paper investigates one structure that plays a significant role in printed dictionaries and is also relevant for e-dictionaries, i.e. the article structure. Lexicographic theory lacks a satisfactory formulation and explanation of this structure in e-dictionaries. A brief overview is given of article structures in printed dictionaries. Then some data from existing e-dictionaries are analysed and discussed. In the discussion suggestions are made for the planning of article structures in e-dictionaries.

2. Article structures in printed dictionaries

2.1 Terminological clarity

In the development of lexicographic theory the German metalexicographer Herbert Ernst Wiegand played a dominant role in research regarding dictionary structures. In numerous publications, cf. Wiegand (1984, 1989, 1989a, 1989b, 1996, 2005), he introduced, identified and discussed a variety of dictionary structures in a meticulous way. Article structures also came to the fore in his publications, cf. Wiegand (2002, 2003) as well as Wiegand, Feinauer and Gouws (2013). The latter contribution forms a basis for the discussion in this paper.

Before discussing article structures, the term *article* or its longer form *dictionary article* has to be understood in an unambiguous way. The term *dictionary*

article has not been used in all lexicographic environments. Quite often, especially in British lexicography, lexicographers have opted for the term *dictionary entry*. This paper uses *dictionary article* as the preferred term because there is an alternative use of the term *dictionary entry*, i.e. as referring to individual parts of dictionary articles and other texts in dictionaries. The default article, albeit not all articles, consists of at least a lemma sign. This lemma sign is the guiding element of the article and is usually followed by further entries presented as part of the treatment of the lemma. According to Wiegand, Feinauer and Gouws (2013: 317) a dictionary article is an accessible dictionary unit and it is characterised by three definitive features, i.e. "(a) It shows at least one external access text element. (b) It is a constituent of a wordlist. (c) It consists of lexicographic data made accessible, including at least one such part, by means of which the dictionary user can unlock lexicographic information mentioned by the access text element."

Accessibility is a critical feature of article structures. In printed dictionaries external data accessibility is established by the outer access structures, cf. Wiegand, Feinauer and Gouws (2013: 325). Another important feature of dictionary articles in printed dictionaries is the fact that they are part of a word list, typically the list containing all the macrostructural elements. This has immediate implications for their possible occurrence in e-dictionaries where word lists are not always an obvious structural component.

2.2 The environment of dictionary articles

In the majority of dictionary types lexicographers employ the macrostructure as an ordering structure to present lemmata as guiding elements of their respective dictionary articles. In a printed dictionary adhering to an alphabetical ordering the main macrostructure will consist of a series of article stretches. In alphabetically ordered dictionaries article stretches typically coincide with the 26 letters of the alphabet. However, this is not always necessarily the case. Some languages do not use specific letters whilst other languages have additional letters. A lexicographer of a dictionary dealing with Northern Sotho could e.g. consider including a separate article stretch for the letter "š", seeing that this letter has a frequent occurrence in this language. Some Afrikaans dictionaries, e.g. *Nasionale woordeboek*, does not have an article stretch "Q" because, based on the needs of the target users of that dictionary, the lemma selection has not yielded any words starting with the letter "q".

According to Wiegand, Beer and Gouws (2013: 43) article stretches are immediate text constituents of the word list of an initial-alphabetical dictionary. As is the case with macrostructures, article structures can also be divided into those with a straight alphabetical ordering and those presented in a horizontal order as niched or nested article stretches. Any subsection of an article stretch, e.g. a given nest or niche, forms a partial article stretch and each partial article stretch is an immediate constituent of a specific article stretch. Wiegand,

Beer and Gouws (2013: 48) make provision for functional partial article stretches, i.e. those partial article stretches "that display a genuine function directed at the dictionary subject matter." Such a functional article stretch could be formed on e.g. morpho-semantic grounds, etymological grounds, semantic relationships, etc. In many monolingual dictionaries that employ a system of horizontal macrostructural ordering the niches or nests are based on their relation with the preceding vertically ordered lemma. Compare in this regard the following excerpt from *Groot woordeboek/Major Dictionary*, a bilingual dictionary with Afrikaans and English as treated language pair, where the article of the lemma sign *dak* is followed by an article niche (of which only a part is given here). This niche is not only a partial article stretch (presenting the articles in the alphabetical stretch *dakbalk-daklose*) but it also is a functional partial article stretch because it includes words with the same lexical item, i.e. *dak-*, as first component and groups them together. With regard to the genuine purpose of functional article stretches Wiegand, Feinauer and Gouws (2013: 328) say: "They further show a genuine purpose, namely to indicate relationships regarding content between the lemma signs treated in the dictionary articles;" This partial article stretch satisfies the genuine function of macrostructural coverage of a specific subsection of the lexicon of Afrikaans.

dak, (-ke), roof; hanging wall (in mines); *op iem. se ~ AFKLIM*, give someone a dressing down; *onder ~ BRING*, give shelter; accommodate; '*n ~ bo jou HOOF hê*, have a roof over one's head; *te veel ~ op die HUIS hê*, walls have ears; *dit KRY ek op my ~*, I shall have to answer for that; *ONDER ~ wees*, have a roof over one's head; *ONDER een ~ sit*, have common interests; *SONDER ~*, roofless; *van die ~ke af VERKONDIG*, proclaim from the housetops; **~balk**, roof beam; **~bedekking**, roofing, roofage; **~bindbalk**, principal beam; **~bint**, roof-tie; **~boog**, carline, hoopstick, roof (bow) stick; **~dekker**, tiler; thatcher; **~drup**, eaves, eaves-drip; **~fees**, wetting the roof; *~fees vier*, wet the roof; **~gesteente**, roof wall; hanging rock; **~geut**, eaves-gutter, eaves; cullis; **~hanger**, roof pendant; roof hanger; **~hoog**, as high as the roof; **~huis**, penthouse; **~juk**, ashlar; **~kamertjie**, attic, garret; **~kap**, roof-truss; roof-pendant; principal; **~kiel** valley; **~lamp**, roof lamp; dome lamp; **~lat**, roof lath; **~leer**, cat ladder; **~lei**, roof(ing) slate; **~lig**, skylight; roof-light; **~loos, (..lose)**, roofless, homeless; **~lose, (-s)**, waif, homeless person;

(*Groot woordeboek/Major Dictionary*)

When looking at articles in e-dictionaries one also needs to look at their environment and at the article stretches to which they belong. This can have implications for the structure of these articles. An article in a niche or nest will often have a different structure compared to articles in a partial article stretch of vertically ordered lemmata. This issue will not be discussed in the present paper.

3. Types of articles

Wiegand (2003) as well as Wiegand, Feinauer and Gouws (2013) argue in favour of a typology of articles. Wiegand, Feinauer and Gouws (2013: 318) distinguish different article formation features, e.g. condensation features, access structural features, macrostructural features, data distribution features, mediostructural features, microstructural features, text architectural features, quantity features, standardisation features, addressing features and search area structural features. These features lead to a typology system, consisting of a variety of typologies and within each one of these typologies a number of different types of articles can be identified.

Albeit that this paper will not repeat the discussion of this classification and the different types of articles identified in the discussion many aspects from that discussion are relevant when it comes to a determination of the structures of dictionary articles. Some of these features will play a role in the present discussion.

Different article types found in printed dictionaries also occur in e-dictionaries, e.g. partially condensed articles, single articles and synopsis articles, complex articles, etc. The use in printed dictionaries of a heterogeneous article structure, even in a single dictionary, is relevant to e-dictionaries. Consequently the notion of comprehensive and restricted articles for e-dictionaries will be introduced in this paper.

4. Structural components of dictionary articles

In terms of Wiegand and Beer (2013: 255, 256) three types of text constituent structures can be distinguished, i.e. article constituent structures and their partial structures, i.e. article microstructures and search area structures. From a slightly different perspective three different types of structural components of articles could be distinguished and will be discussed in the following sections, i.e. *text segments*, *comments* and *search zones*. These components form part of the structures of articles in printed dictionaries but they also play an important part in e-dictionaries. Lexicographers of e-dictionaries will do well to familiarise themselves with these components in order to plan their dictionaries in a much better way.

4.1 Text segments

Wiegand and Smit (2013: 152) regard a functional text segment as a

part of a condensed lexicographical accessive entry consisting of a form and at least one genuine function which is assigned to the form in its entirety.

The nature and functions of text segments need to be evaluated in order to determine their occurrence within the articles of e-dictionaries. Articles contain

two types of text segments, i.e. items and indicators. Items are given as part of the microstructure whereas indicators do not belong to the microstructure but rather to the article structure. Items are data-carrying entries from which the user can retrieve information relevant to fulfilling the genuine purpose of the specific dictionary. These entries include e.g. items giving the pronunciation, morphology, part of speech, paraphrase of meaning, translation equivalents, illustrative examples, etc. Structural indicators are not data-carrying entries but they identify certain items or article slots. Wiegand (1989: 428) distinguishes between typographical, e.g. bold, italics, roman, etc., and non-typographical structural indicators, e.g. symbols like *, ►, → and ■, or other markers used in the dictionary article, including semi-colons and commas or numbers marking the subcomments on semantics in the article of a lemma representing a polysemous lexical item. According to Wiegand and Smit (2013: 153):

A non-typographical microstructural indicator is a functional text segment without item function but with text constituent status within an article constituent structure which denotes microstructural text constituents.

The articles of any e-dictionary will necessarily contain items. The way in which they are presented will be discussed in subsequent paragraphs. The one significant difference between printed dictionaries and many e-dictionaries lies in the use and especially the nature of structural indicators. Items need to be identified and made accessible and this is the role of indicators in printed dictionaries. Some e-dictionaries maintain a more or less traditional use of both items and structural indicators as their text segments. However, although many e-dictionaries maintain some traditional structural indicators, others deviate in many ways from the approach prevalent in printed dictionaries. This constitutes a new feature of the article structure in e-dictionaries that needs to be investigated from a theoretical perspective.

Gouws (2014, in press) indicates that from the point of view of Information Science structural indicators in e-dictionaries are regarded as a part of the metadata because they guide the user to the required data. A type of entry that is also regarded as metadata in the field of Information Science are the entries giving an identification of the data because the user can retrieve information regarding the subject matter of the dictionary from them. They can be functionally segmented and are therefore also part of the rapid access structure. Data-identifying entries, e.g. "COLLOCATION", "ANTONYM", etc. are structural indicators that give an explicit identification of the type of data accommodated in the relevant search zone. Varieties of this type of structural indicators often occur in e-dictionaries.

When designing e-dictionaries this use of indicators should be taken into account. Bothma and Gouws (2013) argue that when designing the article structure the lexicographer should decide on the levels of granularity and which fields need to be named for the sake of clarity. This is where data-identifying entries as a type of structural indicator can play an important role. It is necessary to draw a hierarchical tree structure of categories and sub-categories and

to define relationships between categories. The database needs to be developed to reflect these decisions.

Within the article structure of printed dictionaries provision is also made for procedures of enrichment and expansion. Expansion implies that a given item can be up-, in- or downwardly extended in order to perform an additional treatment. Enrichment implies that a part of an item is typographically adapted, e.g. by means of caps, to distinguish it from other sections of the item and to put special focus on that item segment. Expansion leads to functional segmentation whereas functional segmentation is not possible in the case of enrichment. This aspect of the text segments of dictionary articles should also be discussed when looking at e-dictionaries.

Further reference to items and the use of indicators in e-dictionaries will be given in the section dealing with search zones.

4.2 Comments

According to Wiegand, Feinauer and Gouws (2013: 343) all immediate text constituents of a condensed dictionary article are called *comments*. Comments bring items together within the microstructure that belong together and by doing so they establish order and coherence in the article. A basic dictionary article consists of two comments, i.e. a comment on form and another comment, more often than not a comment on semantics. The following partial article from the *Woordeboek van die Afrikaanse Taal* (the WAT) has a comment on form and a comment on semantics. It shows an article-architecture with the data arranged in various text blocks — the comment on form in the first text block and the comment on semantics consisting of different subcomments on semantics in subsequent text blocks:

¹**maag** s.nw., mae of soms mage en selde, veral in streek- of geselstaal, maagte; magie.

1 Sakagtige spysverteringsorgaan i.d. buikholte b.d. mens, wat die wydste deel v.d. spysverteringskanaal vorm tussen die slukderm en die duodenum, waarin die pas afgeslukte kos versamel, die koolhidrate gedeeltelik verteer word en die vertering van proteïene begin; sin. *buik* (*verhewe* of *verouderend*), *pens* (*plat*), *kossak* (*skertsend*; *ongewoon*): 'n *Leë, vol maag*. 'n *Sterk, gesonde, goeie maag*. 'n *Swak maag*. *Pyn, krampe op jou maag hê*. *Naar op jou maag voel*. *Medisyne op jou nugter maag drink*. *Met 'n honger maag bed toe gaan*. *Moenie op 'n vol maag gaan swem nie*. ≈ 'n *Maag wat gesond is, verteer maklik gewone voedsel* (O. Schild: *Voedingsl.*, s.j., 94). *Hy het verlang om sy maag te vul met die peule wat die varke eet* (Luk. 15: 16, 1953). *Uitgeput ... en met honger mae bereik die famielie ná 'n vermoeiende stap die wa* (Huisg., 27 Sept. 1929, 17). *My kinders (word) nie verwaarloos ... nie ... hulle eet hulle mage trommelvol* (J. v. Bruggen: *Sluis*¹², 1957, 48). *Ons, die*

.....

2 Min of meer soortgelyke orgaan of enigeen van verskeie soortgelyke organe by diere; dikw. sin. met pens: *Herkouende diere het 'n ... saamgestelde maag ... Dit is saamgestel uit vier afdelings* (C.S. Grobbelaar in Pop.-Wet. Leesb. VI, 1920, 7). *Die voedsel (ondergaan)* (by herkouers) *eers 'n voorlopige gisting in die voormaag ... voordat dit die werklike maag of melkpens en dermkanaal bereik*).

.....
3 a i Uitwendige gedeelte v.d. buik (BUIK I 1) v.d. mens; sin. *buik* (*verhewe* of *verouderend*), pens (*plat*); ook, (*minder verhewe*) buik (BUIK I 1): *Plat op jou maag lê. Oor sy maag vryf.*

(Woordeboek van die Afrikaanse Taal)

The articles in printed dictionaries display different comments, combinations of comments and article-internal positioning of comments. This will not be discussed in the present paper. Within the articles of e-dictionaries a similar occurrence of comments as seen in the article from the WAT is often found, cf. the following partial article from the monolingual *Van Dale* dictionary (Figure 1) where an article-architecture also prevails and where the items dealing with the comment on form have been allocated the first text block and the subcomments on semantics, constituting the comment on semantics, are ordered in subsequent text blocks:

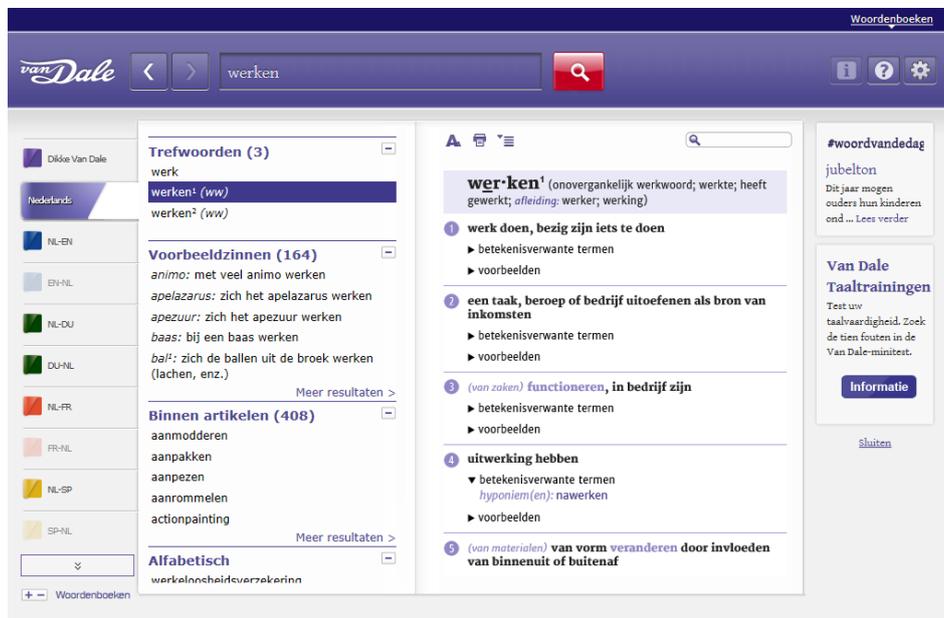


Figure 1

The comments are venues to accommodate different types of lexicographic data. The data distribution in any dictionary needs to take the needs of the target users into account. This could result in a different ordering of data or even of comments, as e.g. seen in *COBUILD* with its innovative centre column, resulting in split comments with some of the typical items belonging to the comment on form given in the traditional position immediately after the lemma sign whilst others, e.g. grammar notes, are allocated a position in the centre column which also accommodates items like notes on semantic relationships from the comment on semantics. This centre column contains a hybrid comment, i.e. a comment on form and semantics.

Lexicographers of e-dictionaries have already employed strategies to deviate from a traditional ordering of comments in their dictionary articles. This is primarily the result of new data distribution options made possible by the e-environment. The dynamic nature of e-dictionaries enables lexicographers to move away from a static to a dynamic data display that includes the use of a multi-layered structure of dictionary articles. This is illustrated by the following screenshots from *lexiko*, one of the innovative online dictionaries of the Institute for German language in Mannheim, Germany. Figure 2 shows the result of a first search for the lemma *Farbe* where the user gets the option to go to the relevant article, cf. Figure 3. This first layer of the article structure, Figure 3, does not show a traditional presentation of a comment on form and a comment on semantics, but rather a rudimentary version of these comments that is divided into two text blocks not displaying the typical distribution of item types. One block contains data, including orthographic data that are usually allocated to the comment on form, but also a data-identifying entry that guides the user to data regarding word formation products. The data in this first text block are relevant to all the senses of the lemma. The second block contains sub-blocks for each of the senses. These sub-blocks contain a paraphrase of meaning and an option to proceed to another layer. Clicking on the structural indicator "weiter" (= further) in anyone of these sub-blocks guides the user to a further layer in the treatment of the lemma, cf. Figure 4. Here the paraphrase of meaning is repeated but the user can move to yet another layer of treatment by clicking on buttons, e.g. to find data on collocations or grammar. Clicking on the data-identifying entry "Grammar" guides a user to the display presented in Figure 5. The grammar data here would typically be part of a traditional comment on form but here it is isolated from the text block containing comment on form items seen in Figure 3.

The *lexiko* article of the lemma *Farbe* contains data comparable to what is found in printed dictionaries where a traditional division between the comment on form and the comment on semantics is maintained. However, a focus on options for rapid access to the specific data regarded as salient by a user in a specific situation of use eschews a structure comprised of a comment on form and a comment on semantics in favour of a multi-layered presentation with rapid access to a different grouping of items and data types that do not adhere to a typical comment on form and comment on semantics division. These excerpts from *lexiko* compel one to identify and acknowledge an important

structure in e-dictionaries, i.e. the screenshot structure, and to realise that the static article structure in printed dictionaries is substituted by a dynamic multi-layered article structure.

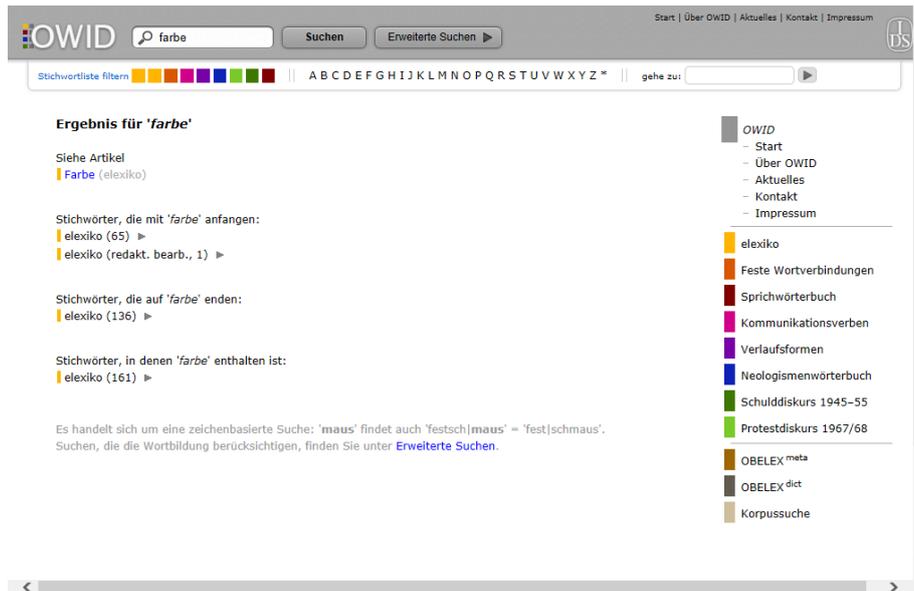


Figure 2

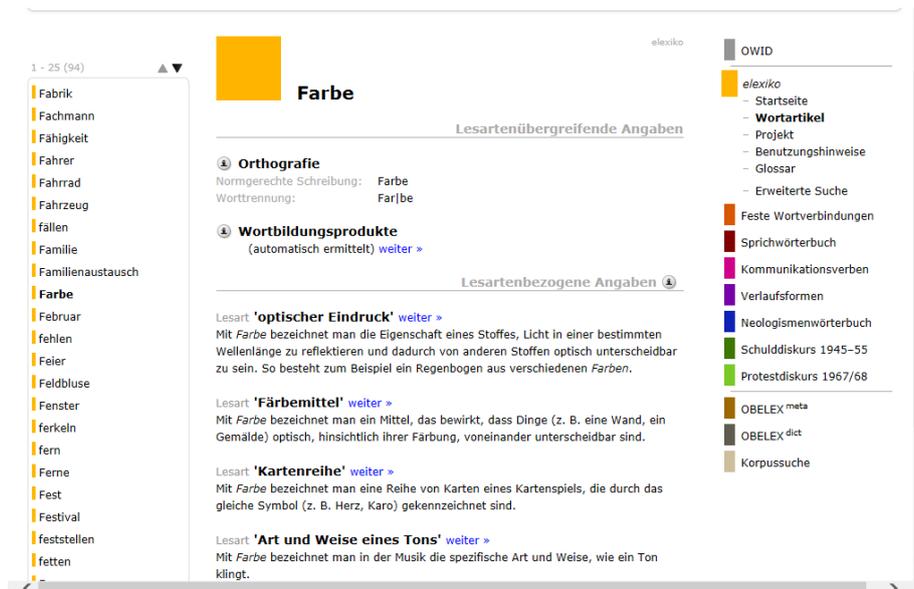


Figure 3

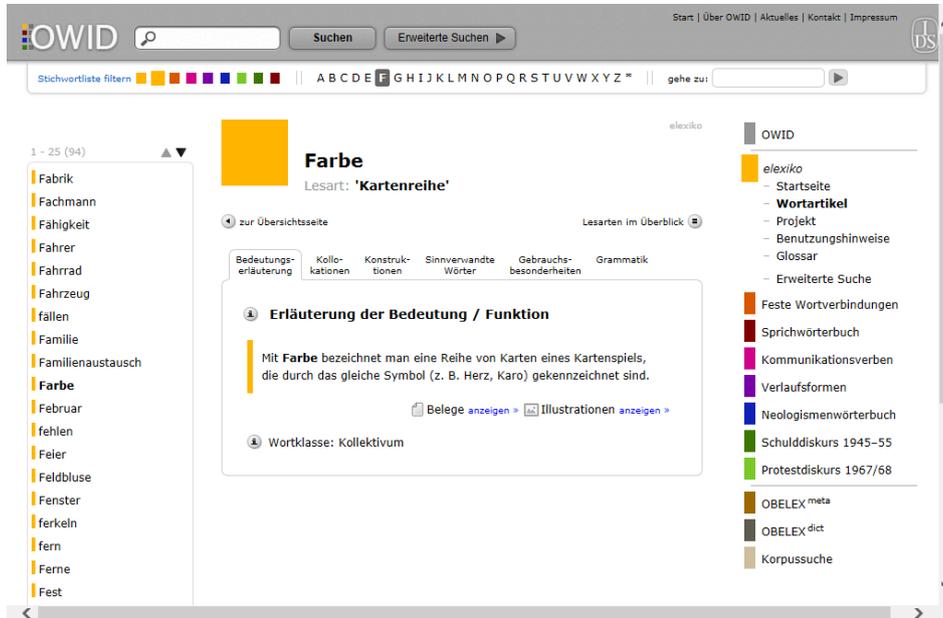


Figure 4

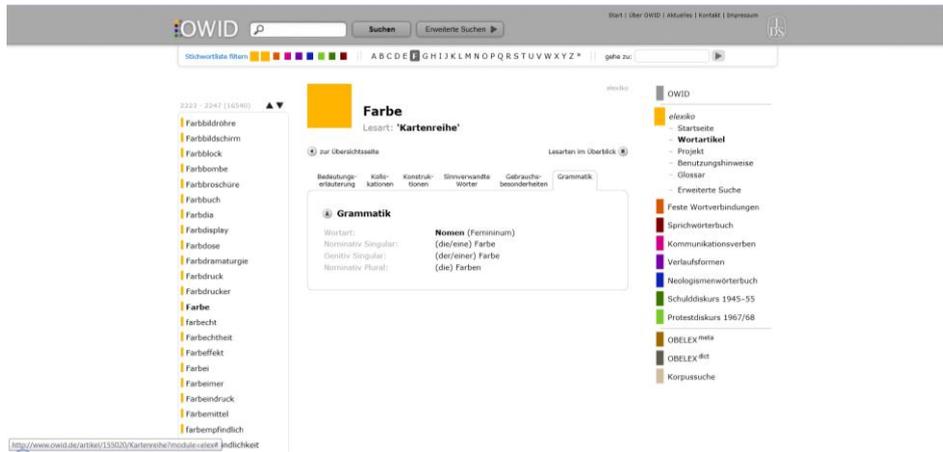


Figure 5

The screenshot structure is not exclusively part of an article structure — it is used in various other positions in e-dictionaries and can e.g. precede access to the article, e.g. as opening shot when consulting an e-dictionary for the first time. But it also plays a vital role in the structure of articles where the screenshot structure initiates access to the search zones, cf. 4.3.

Deviation from an absolute division between the comment on form and the comment on semantics is also seen in the German *Duden Online*. The following three screenshots, given here as three screenshots but actually representing a single display in the dictionary, show how the data in the article of the lemma *Fenster* have been ordered in different sections, each with a formal heading to indicate the data type, e.g. spelling, meaning, Did you know?, synonyms, pronunciation, origin, grammar, combinations. This article exhibits a definite article-architecture but the allocation of items to sections supersedes their allocation to comments. The topics that introduce each section help with explicit and rapid access to the relevant data so that the need for the cohesion provided by an ordering in two distinct comments becomes superfluous. This is also a multi-layered article structure. When clicking e.g. on a given paraphrase of meaning a screenshot opens with illustrative examples of all the different senses of the word *Fenster*, cf. Figure 9.



Figure 6

The article structure in e-dictionaries can accommodate different comments but the new environment favours different procedures of grouping and access routes to data that do not lead via comments as article components. The allocation of lexicographic data to easily accessible search zones is of more importance than a strict adherence to a division in different comments.

Bedeutungen ⓘ Nach oben

- meist verglaste Öffnung, die Licht [und Luft] in einen geschlossenen Raum dringen lässt
 - zum Verschließen der Fensteröffnung dienendes gerahmtes Glas
- (umgangssprachlich) Kurzform für: Schaufenster
- (EDV) auf dem Bildschirm eines Computers erscheinendes rechteckiges Feld, das dazu dient, Textteile, Teile eines Programmes oder andere Informationen zusätzlich auf der Benutzeroberfläche darzustellen [und zu bearbeiten]

Wussten Sie schon? ⓘ Nach oben

Dieses Wort gehört zum Wortschatz des Zertifikats Deutsch.

Synonyme ⓘ **zu Fenster** Nach oben

- Fensteröffnung
- Fensterglas, [Fenster]scheibe, Fensterverglasung, Glasfenster
- Auslage, Ladenfenster, Schaufenster
- (EDV) Window

Aussprache ⓘ Nach oben

Betonung:
Fenster ⓘ

Figure 7

Fenster ⓘ

Herkunft ⓘ Nach oben

mittelhochdeutsch venster, althochdeutsch fenstar < lateinisch fenestra

Grammatik ⓘ Nach oben

	Singular	Plural
Nominativ	das Fenster	die Fenster
Genitiv	des Fensters	der Fenster
Dativ	dem Fenster	den Fenstern
Akkusativ	das Fenster	die Fenster

Typische Verbindungen (computergeneriert) ⓘ Nach oben

Adjektive Verben Substantive

offen geheim klein geschlossen

Figure 8

Bedeutungen, Beispiele und Wendungen  Nach oben

1.

a. meist verglaste Öffnung, die Licht [und Luft] in einen geschlossenen Raum dringen lässt

Beispiele

- ein vergittertes Fenster
- das Fenster geht auf die Straße hinaus
- aus dem Fenster sehen, fallen
- die Nachbarn lagen alle in den Fenstern
(*stützten sich mit den Armen auf die Fensterbank und sahen hinaus*)
- <in übertragener Bedeutung>: ein Briefumschlag mit Fenster (*Fensterbriefumschlag*)



Geöffnetes Fenster - © Kerry - Fololia.com

Wendungen, Redensarten, Sprichwörter

- sich [zu] weit aus dem Fenster lehnen/hängen (sich [zu] weit vorwagen, sich [zu] sehr exponieren)
- aus dem/zum Fenster hinausreden, Fenstersprechen (1. vergeblich, erfolglos reden, sprechen. 2. scheinbar nur um der Sache willen, in Wahrheit aber für die Öffentlichkeit propagandistisch, manipulierend reden, sprechen.)
- weg vom Fenster sein (*umgangssprachlich*; [von der Öffentlichkeit] nicht mehr beachtet sein, abgeschrieben, nicht mehr gefragt sein)

b. zum Verschließen der Fensteröffnung dienendes gerahmtes Glas

Beispiele

- ein geöffnetes, gekipptes, eingeschlagenes Fenster
- das Fenster ist blind geworden
- die Fenster putzen
- <in übertragener Bedeutung>: mit dieser Maßnahme hat der Staat endlich wieder ein

Figure 9

4.3 Search zones

Search zones constitute another category of structural components of articles in printed dictionaries. It also applies to e-dictionaries where adaptations can enhance the value of search zones to ensure a rapid and unambiguous retrieval of information. One of the advantages of a well-devised search zone structure is that it increases the user-friendliness of the dictionary considerably because of the clear demarcation of slots accommodating the different data types. The data distribution and presentation in e-dictionaries needs to be done in a meticulous way and in this regard the planning of clearly identifiable, findable and accessible search zones is of extreme importance. Search zones are subsections of dictionary articles and form part of a more comprehensive data distribution environment. The word list of a printed dictionary forms a search field, each dictionary article is a search area and all these search areas are divided into search zones, cf. Wiegand and Beer (2013). With regard to e-dictionaries it could be argued that the article stretch compound constitutes a search field with the articles and their subdivisions fulfilling a similar role as their counterparts in printed dictionaries. This is illustrated by the following partial article from the *Oxford English Dictionary*, the OED, Figure 10:

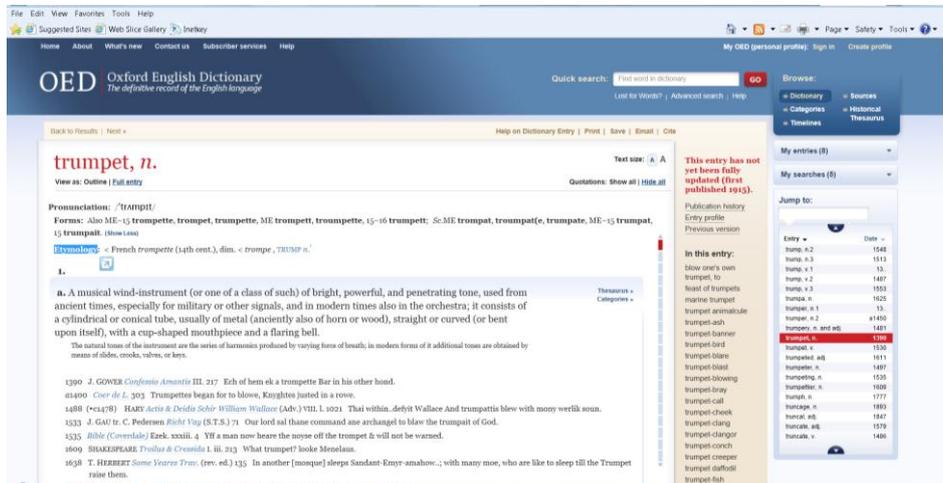


Figure 10

In this article different data types are clearly presented in different text blocks, i.e. search zones.

Wiegand, Beer and Gouws (2013: 64) argue that

Search zones have the general genuine function to assist the knowledgeable active dictionary user by means of search zone indicators in making the search for the items regarded as search targets easier.

They also stress the importance of a classification system of search zone indicators.

As is the case with comments, search zones could also have a similar display in printed and e-dictionaries, e.g. with regard to the use of typographical and non-typographical structural indicators, cf. Figure 11 from the WAT. Here one finds typographical indicators to present the lemma in bold, the paraphrases of meaning in roman and examples in italics and non-typographical indicators like numbers as polysemy markers and the black square "■" marking specific types of additional data.

This similar use of especially non-typographical structural indicators as in printed dictionaries is often found in e-dictionaries that have a printed dictionary as its base. However, in dictionaries designed for the e-medium this use of non-typographical indicators is not needed because a multi-layered presentation of search zones typically employs data-identifying entries to guide the user to the relevant items or search zone. This can be seen in e.g. the data presentation and article structure of the previously discussed *lexiko* excerpts. As previously indicated in the discussion of *lexiko* the full article is not displayed at any one time. However, users can move from one layer to another by clicking on buttons that function as data-identifying entries as seen in Figure 4 from *lexiko*.

The screenshot shows the WAT Elektroniese website interface. At the top, there is a navigation menu with options like 'Tuisblad', 'WAT op selfoon', 'WAT-inligting', 'Kontakinligting', 'Bestelvorms', and 'Registrasievorm'. Below this is a search bar with the text 'Huidige soektog: rooi Resultate: 2620'. A list of related words is shown on the left, including 'rooifend', 'rooibou', 'rooifler', 'rooifluiker', 'rooifhandel', 'rooifluweik', 'rooifyk', 'rooifyker', 'rooifluster', 'rooiflusteraar', 'rooifmeu', 'rooifridder', 'rooifskip', 'rooifug', 'rooifaxo', 'rooifog', 'rooifvangs', 'rooifvis', 'rooifvisser', 'rooifvisserij', 'rooifvlieg', 'rooivoel', 'rooifvyand', 'rooi', 'rooialbessie', 'rooiaarde', 'rooias', 'rooiasdiertjie', 'rooifrikaner', 'rooigurke', 'rooisappeliefie', and 'rooissie'. The main content area displays the definition of 'rooi' and several example sentences with the word highlighted in yellow.

Figure 11

These data-identifying entries, e.g. "Kollokationen", "Konstruktionen" and "Grammatik", have the same function as structural indicators in printed dictionaries. They are the typical non-typographical structural indicators in e-dictionaries. Because they guide the user to the relevant search zone the use of other non-typographical indicators like symbols is not needed, cf. Bothma and Gouws (2013). These data-identifying entries are much more explicit, cf. Gouws (2014, in press), and users do not have to learn the value of a range of symbols that are used as non-typographical indicators.

In the articles of e-dictionaries the use of clearly-identified search zones is of extreme importance. Their occurrence in a multi-layered presentation within the article as search area could be regarded as one of the innovative aspects of e-dictionaries for which further research, also with regard to the use of data-identifying entries as non-typographical structural indicators, can only enhance the lexicographic endeavour. In terms of Wiegand (2005) these non-typographical indicators need to help users with their article-internal orientation so that they are in a position to identify, distinguish and group different items.

5. Different articles with the same guiding element

It has already been shown that some e-dictionaries display an article structure that resembles that of the printed version of the dictionary, cf. Figure 12 from the *American Heritage Dictionary*:



Figure 12

Lexicographers of e-dictionaries do have the option to structure articles in this way but in the planning of an online dictionary one needs to negotiate the spectrum of possibilities that the medium offers. One of the major advantages of online dictionaries is that users do not necessarily have to work through all search zones or items a dictionary offers in its default articles. Embarking on the e-possibilities some online dictionaries give access to the full article but, as seen in 4.3, also allow the user to go directly to a specific search zone or the item giving the data he/she requires. This is done by means of rapid access using data-identifying entries that can either guide the user to a specific item or to a selected search zone. These procedures typically result in screenshots where the lemma is presented, complemented only by the desired data type, e.g. only an item giving the paraphrase of meaning, or only the translation equivalents or only whatever data type the user is looking for.

The entries displayed in such a presentation can also function as part of the comprehensive article structure of the given lemma but for the specific consultation such a lemma plus a single search zone or a single item function as fully-fledged articles. When consulting the Dutch *Algemeen Nederlands Woordenboek* the user may only be interested in the meaning of the word *fado*. Utilising the search options he/she may then reach the following destination, cf. Figure 13, i.e. a rudimentary dictionary article with *fado* as lemma, complemented by two abbreviated paraphrases of meaning to identify the different senses of this word.



Figure 13

If the user is interested in the first polysemous sense he/she clicks on that sense, as data-identifying entry, and is guided to an article with a more comprehensive treatment directed at *fado* in that specific sense:

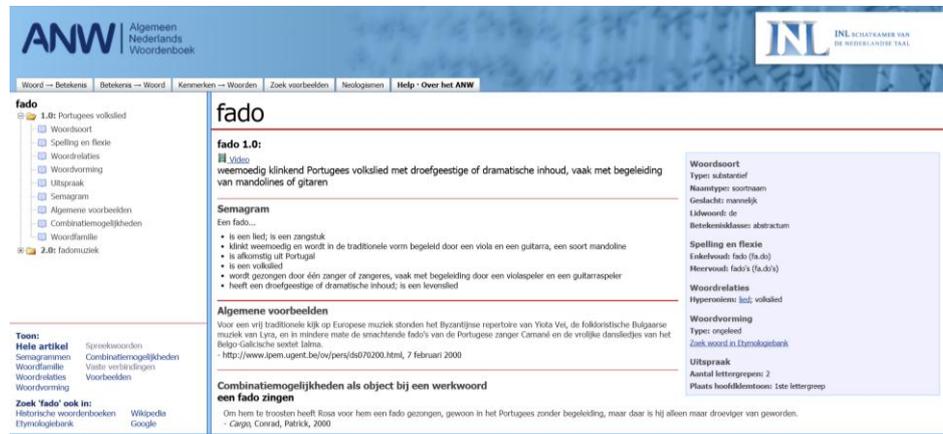


Figure 14

Interest in one data type, e.g. "word family" leads by means of a single click to the following article:

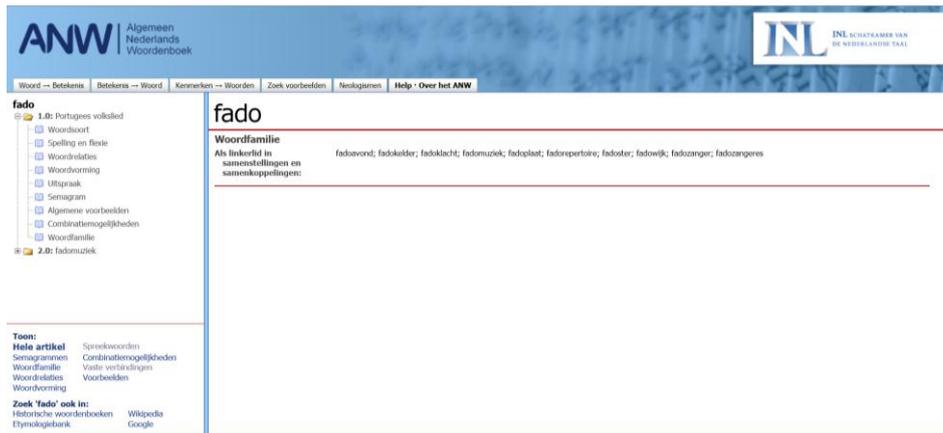


Figure 15

A click on the entry *Hele artikel* (= full article) gives access to the full article. The article structure with its layered search zones and entries guiding the user to specific items meet the important criterion of external data-accessibility, as formulated by Wiegand, Feinauer and Gouws (2013: 325).

An access process that can direct searches to both comprehensive articles and articles with a restricted structure that are in complete accordance with the needs of the user elevates the value of a dictionary. When planning the treatment allocated to lemmata in e-dictionaries lexicographers should consider the possibility of layering search zones in such a way that the user can find various articles, each with its own unique article structure, for a single lemma where, once again, the function and the needs of the user determine the article structure, contents and presentation of data. The user should be able to extract different articles that have the same lemma from one single database and to access those articles and the specific data type by means of a search option determined by the specific need of the user in a specific situation of use.

Even within the domains of one function different articles with the same lemma should be possible. If the dictionary e.g. has a cognitive function the user should be able to access different data types included as part of the cognitive task identified for that dictionary. This implies that if the user is in need of etymological data a consultation should lead him/her directly to the etymological data and not to e.g. some encyclopaedic data in the paraphrase of meaning that have been included as part of the response to the cognitive function of the dictionary. Search zones are designed in terms of data types and not functions. However, for future dictionaries lexicographers could consider access to function-based search zones in the articles of bi- or polyfunctional dictionaries.

Different articles do not only have to be pre-planned in terms of their structure and the types of items to be included. These pre-planned articles, i.e. either comprehensive or restricted in format, can be seen as part of the default variation in structure. However, a further type of variation in article structure is needed in an online dictionary. The user should be given the opportunity to either use the default settings of the dictionary, i.e. e.g. the option to use the pre-planned and pre-set article structures as indicated above, or to create a user profile according to which any search will lead to the retrieval of information directly responding to the specific needs of the user in terms of the given profile. This implies that for a given user who has created his/her profile a niche-made article can be retrieved that is unique for that specific consultation procedure and the specific needs of the active user.

Within each dictionary easy and direct access to the individual data types or to the article as a whole should result from a choice the user can make at an early phase of the dictionary consultation process. The setting up of a user profile should also allow the establishment of the default article type for the specific user during a comparable consultation process. A prerequisite for these possibilities is a well-designed system of article structures that have been developed

in accordance with the underlying theoretical basis of the dictionary.

Planning of a new online dictionary could also be done in such a way that not only different articles of the same lemma can be accessed but from a single database users should actually be able to access different dictionaries treating that lemma, cf. Bergenholtz and Nielsen (2013: 82). This implies a proper marking-up of database entries so that the retrieval of information is in line with the specific dictionary and the needs of the user in terms of his/her selected profile. The items needed for the different dictionaries need to be accommodated in a specific microstructure which is a component of the article structure. Different types of microstructures need to be considered for each dictionary in order to get the best structure for the ordering of the items in the specific articles. This should also be done with regard to the functions, users and situation of use of the specific dictionary.

6. Conclusion

The importance of lexicographic structures in e-dictionaries should not be underestimated but lexicographers may not follow an uncritical approach by merely trying to utilise the same structures as employed in printed dictionaries. The article structure needs to be maintained in e-dictionaries but with a different focus. As immediate constituents of article and partial article stretches, cf. Gouws (2014a, in press), or as immediate results of lexicographic search processes not going via an article stretch or an alphabet bar the division of an article into two or more comments should not be seen as a default structure. Data should rather be retrieved from different search zones constituting a multi-layered article structure with a variety of screen shots to present the relevant lexicographic data. Articles will always contain both items and structural indicators as text segments but the structural indicators in e-dictionaries will rather be of a non-typographic nature, utilising more data-identifying entries to ensure unambiguous and rapid access to the relevant search zones and items.

Access to specific search zones and data types with a required search zone being the only article component to complement the lemma sign leads to the acknowledgement of restricted articles. The data in these articles can also be presented in the comprehensive version of the article but its occurrence as only complement to the lemma qualifies the presentation as a fully-fledged article in its own right. This leads to heterogeneous article structures in a variety of restricted articles that share a single lemma. This lemma also functions as guiding element of a comprehensive article.

In the planning and compilation of e-dictionaries a varied article structure and the possibility to retrieve different articles with the same lemma from a single data base represent innovative features that do not prevail in printed dictionaries. The data distribution allocating items to hierarchically-ordered multi-layered search zones needs to be planned in accordance with the user needs and specific situations of use in order to enhance the success of the dic-

tionary consultation process.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85434).

The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.

7. Bibliography

7.1 Dictionaries

7.1.1 Printed dictionaries

Botha, W.F. (Ed.-in-Chief). 1996. *Woordeboek van die Afrikaanse Taal, Volume X*. Stellenbosch: Bureau of the WAT.

De Villiers, Meyer et al. (Eds.). 1987⁷. *Nasionale Woordeboek*. Cape Town: Nasou.

Eksteun, L.C. et al. (Eds.). 1997¹⁴. *Groot Woordeboek Afrikaans–Engels/Engels–Afrikaans / Major Dictionary Afrikaans–English/English–Afrikaans*. Cape Town: Pharos.

Sinclair, J.M. (Ed.). 1987. *Collins COBUILD English Language Dictionary*. London/Glasgow: Harper-Collins.

7.1.2 Online dictionaries

Algemeen Nederlands Woordenboek. <http://anw.inl.nl/search>

American Heritage Dictionary. <http://ahdictionary.com/>

Duden Online. <http://www.duden.de/>

elexiko. <http://www.owid.de/wb/elexiko/start.html>

Oxford English Dictionary. <http://www.oed.com/>

Van Dale woordenboeken. <http://www.vandale.be/opzoeken>

Woordeboek van die Afrikaanse Taal. <http://www.woordeboek.co.za.ez.sun.ac.za/>

7.2 Other literature

Bergenholtz, H. and S. Nielsen. 2013. What is a Lexicographical Database? *Lexikos* 23: 77-87.

Bothma, T.J.D. and R.H. Gouws. 2013. *Mapping Indicators in Print Dictionaries to Structural Descriptions in e-Dictionaries*. Unpublished paper presented at the 2013 eLex Conference, 17–19 October 2013, Tallinn.

- Gouws, R.H.** 2014. In press. *Aspekte des lexikographischen Prozesses in Print- und Onlinewörterbüchern. OPAL — Online publizierte Arbeiten zur Linguistik*. Mannheim: Institut für Deutsche Sprache.
- Gouws, R.H.** 2014a. In press. *Makrostruktuuraanpassings vanaf gedrukte na e-woordeboeke*.
- Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.)**. 2013. Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/ New York: Walter de Gruyter.
- Hausmann, F.J. et al. (Eds.)**. 1989–1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*. Berlin: Walter de Gruyter.
- Wiegand, H.E.** 1984. On the Structure and Contents of a General Theory of Lexicography. Hartmann, R.R.K. (Ed.). 1984. *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983*: 13-30. Lexicographica. Series Maior 1. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. Hausmann, F.J. et al. (Eds.). 1989–1991: 409-462.
- Wiegand, H.E.** 1989a. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. (Eds.). 1989–1991: 371-409.
- Wiegand, H.E.** 1989b. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Eds.). 1989–1991: 462-501.
- 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 70. Tübingen: Max Niemeyer
- Wiegand, H.E.** 2002. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im *de Gruyter Wörterbuch Deutsch als Fremdsprache*. Zugleich ein Betrag zur Weiterentwicklung einer Theorie der Wörterbuchform. Wiegand, H.E. (Ed.). 2002. *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des de Gruyter Wörterbuchs Deutsch als Fremdsprache*: 497-595. Lexicographica. Series Maior 110. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 2003. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Betrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wiegand, H.E.** 2005. Angaben, funktionale Angabezusätze, Angabetexte, Angabestrukturen, Strukturanzeiger, Kommentare und mehr. Ein Betrag zur Theorie der Wörterbuchform. *Lexicographica. International Annual for Lexicography* 21: 202-379.
- Wiegand, H.E. and S. Beer.** 2013. Access Structures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 110-149.
- Wiegand, H.E., S. Beer and R.H. Gouws.** 2013. Textual Structures in Printed Dictionaries: An Overview. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 31-73.
- Wiegand, H.E., I. Feinauer and R.H. Gouws.** 2013. Types of Dictionary Articles in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 314-366.
- Wiegand, H.E. and M. Smit.** 2013. Microstructures in Printed Dictionaries. Gouws, Rufus H., Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Eds.). 2013: 149-214.

Towards a Southern African English Defining Vocabulary*

Lorna Hiles, *Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa (lorna@dictionarian.co.za)*

Abstract: Controlled defining vocabularies have been used regularly in lexicography since the 1970s. They are mostly employed in learners' and school dictionaries and have been used to ensure that the definitions are easily understood by the dictionary user, and that the words used to define a lemma are not more difficult than the lemma itself.

In my experience, defining vocabularies compiled for English dictionaries for a British or American market are not entirely sufficient for southern African dictionaries. Words which would be well-known and used in southern Africa are not included in a foreign defining vocabulary.

With this article, it is my intention to find methods of adapting existing defining vocabularies and to compile one that is more useful for southern African school dictionaries.

A complete southern African English defining vocabulary is not part of the scope of this paper, but by providing some suggestions, I hope to make the reader realize that there is a need for a complete defining vocabulary that would cater to the needs of a southern African dictionary and its users. Further research would then be an extension of this paper into a full defining vocabulary (DV) that would be used for southern African dictionaries.

Keywords: CONTROLLED DEFINING VOCABULARY, DEFINING VOCABULARY, DEFINITIONS, LEARNERS' DICTIONARY, RESTRICTED DEFINING VOCABULARY, SCHOOL DICTIONARY, SOUTH AFRICAN DICTIONARIES, SOUTHERN AFRICAN DICTIONARIES

Opsomming: Op pad na 'n Suider-Afrikaanse Engelse definiëringswoorde-skat. 'n Gekontroleerde definiëringswoorde-skat word gereeld sedert die 1970's in die leksikografie gebruik. Dit word meesal in aanleerders- en skoolwoordeboeke benut en word gebruik om seker te maak dat die omskrywings geredelik verstaanbaar is vir die woordeboekgebruiker, en dat die woorde wat gebruik word om 'n lemma te omskryf, nie moeiliker is as die lemma self nie.

Die definiëringswoorde-skat wat vir Engelse woordeboeke vir Britse of Amerikaanse markte saamgestel word, is nie heeltemal voldoende vir Suider-Afrikaanse woordeboeke nie. Woorde wat in Suider-Afrika goed bekend is en gebruik word, word nie by 'n oorsese definiëringswoorde-skat ingesluit nie.

Met hierdie artikel is dit die voorneme om metodes te vind om die bestaande definiëringswoorde-skat aan te pas en om een saam te stel wat nuttiger is vir Suider-Afrikaanse skoolwoordeboeke.

'n Omvattende Suider-Afrikaanse definiëringswoorde-skat is nie deel van die bestek van hier-

* This article was presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

die artikel nie. Deur 'n paar voorstelle te maak, behoort die leser daarvan bewus gemaak te word dat daar 'n omvattende definiëringswoordeboek nodig is wat in die behoeftes van 'n Suider-Afrikaanse woordeboek en die gebruikers daarvan voorsien. Hierdie artikel sou as wegspringplek kon dien vir voortgesette navorsing na 'n definiëringswoordeboek wat vir Suider-Afrikaanse woordeboeke van nut kan wees.

Sleutelwoorde: AANLEERDERSWOORDEBOEK, BEPERKTE DEFINIËRINGSWOORDESKAT, DEFINIËRINGSWOORDESKAT, GEKONTROLEERDE DEFINIËRINGSWOORDESKAT, OMSKRYWINGS, SKOOLWOORDEBOEK, SUID-AFRIKAANSE WOORDEBOEKE, SUIDER-AFRIKAANSE WOORDEBOEKE

Introduction

Restricted defining vocabularies are used in pedagogical dictionaries — these being school dictionaries or learners' dictionaries. They are used to ensure that learners are able to understand the dictionary definitions with their vocabulary level. According to Herbst (1996: 324) a "controlled defining vocabulary is, quite obviously, used to avoid using words in the definitions which are less accessible to the learner than the word being defined". Defining vocabularies are mostly used for definitions, but also offer guidelines for examples and notes.

Defining vocabularies are developed and owned, for the most part, by commercial publishers. They usually contain anything from 2 000 to 5 000 words. They are generally compiled using corpus frequency data, which assumes that the most frequent words in a language are the ones that learners are taught first. "The lexicographer needs words of high generality, and some of these, such as *house*, are also very frequent, but others, such as *vessel*, are not." (Whitcut 1988: 49) The compiler then systematically works through a set of parameters, such as avoiding synonyms and antonyms, to determine which words are necessary to write definitions in a concise and simple way.

It has been found that existing defining vocabularies lack certain words that would make definitions more accessible to southern African learners, and therefore there is a need for a defining vocabulary that would be more suitable for southern African dictionaries — with regard to both their definitions and their example sentences. So, for this paper, methods that can be used to build a Southern African English defining vocabulary have been established. Two methods of examining the existing defining vocabularies have been used to determine what words can be deleted and what words should be added. A method of compiling a southern African defining vocabulary from scratch has also been suggested.

Defining vocabularies

The following existing defining vocabularies were worked with:

- Oxford South Africa, which contains 2672 entries. It was adapted in 2008

for the *South African Oxford School Dictionary 3e* and has been used with permission.

- Oxford 3000, which contains 3 540 entries, and is available on the Internet.
- Macmillan Dictionary Defining Vocabulary, which contains 2450 entries and is available on the Internet.
- Longman American Defining Vocabulary, which contains 2203 entries and is available on the Internet.

In the information about these defining vocabularies, the following comments are made: "The Defining Vocabulary has been carefully chosen after a thorough study of all the well-known frequency lists of English words. Furthermore, only the most common and 'central' meanings of the words on the list have actually been used in definitions." (Longman Dictionaries USA, 2014)

The Macmillan defining vocabulary, "... includes a few simple phrasal verbs like *put on* and *take off* (for talking about clothes), and a few simple phrases like *in order to* and *pay attention*. If they are not listed, we do not use them." (Macmillan Dictionary, 2009)

This research attempts to determine how much needs to change to make these existing defining vocabularies more useful for southern African lexicographers. These changes could be applied globally, but this research focuses specifically on the needs of southern African dictionaries.

Methods of examining defining vocabularies

The first method used in examining these defining vocabularies involved taking a random stretch of words from each of the defining vocabularies and combining them to compile a "megalist" of all the words in that particular stretch. See Appendix A. This megalist was used to search through a South African school dictionary from a different publisher. This dictionary does not use a defining vocabulary, so it was used to establish which of the DV words in the megalist are used in the unrestricted definitions.

The stretch used was from *lightning* to *low*. The megalist contained 62 words, while Oxford 3000 contained 53, Oxford South Africa 42, Macmillan 42, and Longman 34.

The Longman stretch contained only words, no phrases. The Macmillan stretch contained words and two phrases, *look after*, *look for*.

The Oxford 3000 stretch contained more phrases, *look after*, *look at*, *look for*, *look forward to*; derivatives such as, *like*, *unlike*, *likely*, *unlikely*, *load*, *unload*; and gave the part of speech where necessary, for example, *like*: *preposition, verb, conjunction*. It also provided International Phonetic Alphabet (IPA) pronunciation guides where necessary, for example at *live* /laiv/ and *live* /lɪv/.

The Oxford South African stretch contained all of the above, except that pronunciation guides were given as phonetic respelling, not as phonetic tran-

scriptions, and the phrase *look something up* was included as well.

Of the 62 words in the megalist, 18 were not used in any of the definitions in the dictionary. Examples of these words are *lodging, lodgings, locate, local, loneliness*.

This was an interesting exercise in comparison, but it would not be very useful as the only method used to compile a southern African defining vocabulary. For a start, one would need to check whether these words are used in many pedagogical dictionaries, not just one. The range of dictionaries would need to include different user groups, different age groups and different levels of English proficiency. Bilingual dictionaries could also be used, with the example sentences interrogated instead of definitions.

This method would, however, make a useful clean-up exercise, allowing one to take out words that are not used in any of the dictionary definitions, or words that are out dated.

The second method used was less systematic, but it was more useful. A selection of words was checked against the definitions in the school dictionary without a defining vocabulary, and then against the defining vocabularies. Words that were in the school dictionary but not in all four of the defining vocabularies were collated.

Lexical sets were mostly adhered to — for example, animals, parts of the body, technical words, foods, school words and irregular plurals, but the selection of words was otherwise random, and according to the author's intuition.

This process presented some interesting results.

One thing that became apparent was a lack of consistency within lexical sets: for example, *herb, spice, spicy* and *flavour* were all in a different number of the defining vocabularies.

Moreover irregular noun forms were not included in all the defining vocabularies, for example, *mice, feet, teeth, children, men, and women*. *Mice, feet* and *teeth* are certainly different enough from the singular that they cannot just be assumed. The convention for defining vocabularies is that "all normal forms of nouns, verbs, and adjectives" can be used in definitions (Macmillan Dictionary, 2009) so it allows for plurals, both regular and irregular, to be assumed. However, irregular plurals and verb forms are not predictable based on their base forms.

One needs to bear in mind that the defining vocabulary is a list of words that the user *should* know. Therefore irregular plurals should be considered for inclusion in a defining vocabulary.

In table 1, one can see that the irregular plurals mentioned above are in zero, one or three of the defining vocabularies. The third column shows how many times they appear in definitions.

School-related terms (Table 2) and technical terms (Table 3) that do not appear in all four defining vocabularies suggest that these defining vocabularies are out of date. The inclusion of modern school-related terms is an indication of the relevance of the defining vocabulary. In South Africa, the curriculum was given an overhaul in the early 2000s and terms such as "pupil" and

"standard" were dropped in favour of "learner" and "grade". It has not been possible to establish when two of the three foreign DVs were most recently updated. The Oxford 3000 was updated at the time of compilation of the Oxford Advanced Learner's Dictionary 6th edition, which was published in 2000.

	number of DV	number of definitions
mice	0	1
men	0	23
women	0	37
children	3	90
feet	0	43
teeth	1	38

Table 1

	number of DV	number of definitions
classroom	2	2
learner	1	3

Table 2

	number of DV	number of definitions
email	3	8
Internet	3	40
software	2	5

Table 3

This process would also be more useful if one checked the words against more southern African dictionaries. Again, the range of dictionaries used would be important. This method is more useful at providing or suggesting words to be added to a southern African defining vocabulary. See Appendix B for a list of words that are used in definitions in one South African school dictionary but do not appear in all four defining vocabularies.

The third method one can use to build a southern African defining vocabulary would be to examine a corpus of southern African English, and then compile the defining vocabulary from scratch. One could then use the two methods detailed above to delete unnecessary words, or add words that are not

necessarily frequent but that are required.

Conclusion

To conclude, existing defining vocabularies can be cleaned up a lot in order to compile an up-to-date southern African English defining vocabulary, and using a range of dictionaries as checks would be a valuable exercise. As mentioned above, both methods used to adapt existing defining vocabularies can be used more extensively and checked against more dictionaries to make them more accurate tools. One could also extend the search to example sentences and try to establish whether the words used in the examples are words that the typical user of a particular dictionary is likely to know.

Bibliography

- Herbst, T.** 1996. On the Way to the Perfect Learners' Dictionary: A First Comparison of OALD5, LDOCE3, COBUILD2 and CIDE. *International Journal of Lexicography* 9(4): 321-357.
- Longman Dictionaries USA. 2014. *The Longman American Defining Vocabulary* [Online] Available from: http://www.longmandictionariesusa.com/longman/defining_vocabulary [Accessed 30 June 2014].
- Macmillan Dictionary. 2009. *Macmillan Dictionary: Clear definitions* [Online] Available from: <http://www.macmillandictionary.com/learn/clear-definitions.html> [Accessed 30 June 2014].
- Whitcut, J.** 1988. Lexicography in Simple Language. *International Journal of Lexicography* 1(1): 49-55.

Appendix A: Megalist of words in one stretch taken from four existing defining vocabularies

lightning	literary	local	long	lost
like	literature	locally adv.	look	lot
likely	litre	locate v.	look after	loud
limb	little	located adj.	look at	loudly
limit	a little det., pron.	location n.	look for	loudness
limited	live /la?v/ adj., adv.	lock v., n.	look forward to	love n., v.
line	live /l?v/ v.	lodging	look sth up	lovely adj.
link	living adj.	lodgings	loose adj.	lover n.
lion	lively	log	loosely adv.	low adj., adv.
lip	liver	logic n.	lord n.	
liquid	load	logical adj.	lorry n. (BrE)	
list	loaf	loneliness	lose	
listen	loan	lonely	loss	

Appendix B: Words that are used in definitions but do not appear in all four defining vocabularies

	Number of DV	number of definitions
adventure	3	2
bucket	1	2
children	3	90
classroom	2	2
container	3	115
disease	3	66
email	3	8
feet	0	43
flavour	3	4
frog	0	2
goat	3	9
herb	0	7
Internet	3	40

learner	1	3
men	0	23
mice	0	1
permanent	3	8
rabbit	2	4
rat	2	1
sauce	3	12
shellfish	1	3
shiny	3	29
sickness	0	3
software	2	5
specific	2	1
spice	2	2
spicy	1	9
spider	2	4
teeth	1	38
temporary	3	5
thunder	1	2
tribe	2	6
women	0	37
worm	1	1
zoo	0	2

The Lemmatization of Loan Words in the *isiNdebele–English* *isiHlathululi-imagama/Dictionary* and Their Successful Incorporation into the Language*

Sponono Mahlangu, *IsiNdebele National Lexicography Unit,
University of Pretoria, Pretoria, South Africa* (sponono.mahlangu@up.ac.za)

Abstract: Researchers in linguistic and lexicographic fields such as Nkondo (1987: 70) and Kamwangamalu (1997: 89) assert that no language is lexically self-sufficient. According to Jafta (1987: 127), the reason for this is because no perfectly homogenous language group exist. There is no living language that can survive without supplementing its vocabulary through borrowing from another or other language(s). Thus Aitchison (2001: 249) is of the view that language gradually transforms itself and it cannot remain unaltered. On the other hand the so called 'purists' disapprove of such alterations because they view these transformations as a process of language corruption because they prefer coining which Mojela (2010: 702) termed indirect borrowing. This article proves and illustrates this notion as correct, especially in the case of African languages. The technical terms and vocabulary in social interaction is based on adjacent South African languages such as English and Afrikaans, which as official languages have inevitably cross-pollinated each other. Researchers also agree that one way of enriching language is through borrowing. The aim of this article is to show that borrowing does not only enrich a language, but it also causes language dilution especially in the case of languages that are less technologically advanced or do not yet function as official languages.

Keywords: BORROWING, CORPUS, DICTIONARY COMPILATION, LEMMATIZATION, LEXICOGRAPHER, LEXICOGRAPHY, ISINDEBELE WORDS, LOAN WORDS, BILINGUAL DICTIONARY

Opsomming: Lemmatisering van leenwoorde in die *isiNdebele–English isiHlathululi-imagama/Dictionary* en die suksesvolle opname daarvan in die taal. Navorsers op linguistiese en leksikografiese terreine soos Nkondo (1987: 70) en Kamwangamalu (1997: 89) voer aan dat geen taal leksikaal selfgenoegsaam is nie. Volgens Jafta (1987: 127) is

* This article is a revised version of a paper presented 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, South Africa, 5–7 July 2006.

die rede hiervoor dat geen perfekte homogene taalgroep bestaan nie. Daar is geen lewende taal wat kan oorleef sonder om sy woordeskat aan te vul deur ontlening uit 'n ander taal of tale nie. Daarom huldig Aitchison (2001: 249) die siening dat taal geleidelik omvorm word en nie onveranderd kan bly nie. Aan die ander kant keur sogenaamde 'puriste' sulke veranderinge af omdat hulle hierdie transformasies as 'n proses van taalkorrupsie beskou aangesien hulle verkies om indirekte ontlenings te vorm, soos wat Mojela (2010: 702) dit noem. Hierdie artikel bevestig en illustreer hierdie aanname as korrek, veral in die geval van Afrikatale. Die tegniese terme en woordeskat in sosiale interaksie is gebaseer op naasliggende Suid-Afrikaanse tale soos Engels en Afrikaans, wat mekaar as amptelike tale onafwendbaar kruisbestuif het. Navorsers is dit ook eens dat een manier om taal te verryk deur ontlening geskied. Die doel van die artikel is om aan te toon dat ontlening nie net 'n taal verryk nie, maar dat dit ook taalverarming verhaas veral in die geval van tale wat minder tegnologies gevorderd of nog nie behoorlik as amptelike tale funksioneer nie.

Sleutelwoorde: ONTLENING, KORPUS, WOORDEBOEKSAMESTELLING, LEMMA-VORMING, LEKSIKOGRAAF, LEKSIKOGRAFIE, ISINDEBELE WOORDE, LEENWOORDE, TWEETALIGE WOORDEBOEK

1. Introduction

IsiNdebele as a standard language was recognized in South Africa in the 1980s. In 1985, it was formalized and introduced into schools. Its first orthography and spelling rules were published in 1982 and again revised in 1995 and 2005.

Prior to 1994 therefore, isiNdebele was one of the languages that were regarded as being marginalised. To ensure the equitable use and development of these languages, the new government enacted the Pan South African Language Act and established a Board to, among others; provide conducive environments for these languages to expedite their development. Afrikaans and English were used as yardstick (<http://www.pansalb.org.za/pansalbhhistory.html>). However, by then, these languages were unequal because they had been separately and independently developed at their particular homeland levels, their development history therefore not being on a par with that of, for instance, Afrikaans.

The Pan South African Language Board's reorganization and renormalization of all South African official languages, followed by the establishment of dictionary institutions in 1999, came at a time when isiNdebele scarcely had anything that could be called a dictionary. All that existed then was *IsiHlathululi-mezwi sesiNdebele* compiled by Shabangu and Swanepoel (1989) that differed only slightly from *IsiNdebele Terminology and Orthography No. 1* (1983). IsiNdebele therefore had little previous experience as far as dictionary-making was concerned and consequently no sound lexicographic contributions could be made to the lexicographic debates. Thus, Skhosana (1999: 222) confirmed that isiNdebele had an insufficient corpus, which is developed from written materials only and that no consultation with language speakers had taken place.

Nine National Lexicography Units were established and registered in

2001, their task being to compile monolingual explanatory dictionaries and other products to help with language development (<http://www.pansalb.org.za/pansalbhistry.html>). One of these was the isiNdebele National Lexicography Unit (known as iZiko lesiHlathululi-mezwi sesiNdebele). The Lexicography Units are based at tertiary institutions throughout South Africa, the IsiNdebele Lexicography Unit being hosted at the University of Pretoria. In 2006, this Unit published its first bilingual dictionary, the *IsiNdebele-English/English-IsiNdebele IsiHlathululi-mezwi/Bilingual and Explanatory Dictionary* (also known as the *IsiNdebele-English IsiHlathululi-magama/Dictionary*, and henceforth referred to as the *IsiNdebele-English Dictionary*). This is a scholar's dictionary containing 15 000 lemmas in total. The experience gained during the compilation of this dictionary proved the need to re-explore some of the latest lexicographic views, because

- (a) unlike other languages that had their specific areas of development (e.g. isiZulu in KwaZulu-Natal, Setswana in the North-West, etc.), isiNdebele is one of the languages that has been intermingling with or spoken within the boundaries of other languages, and
- (b) isiNdebele had not been recognized as an official, autonomous language until 1985.

2. Language contact influencing isiNdebele

It has been noticed that most of the isiNdebele books from which the 2.1 million running words in the isiNdebele electronic corpus were drawn, have used more borrowed words than indigenous isiNdebele words. Thus a part of isiNdebele's vocabulary has gradually been disposed of and replaced by new Ndebelized foreign words, mainly from Afrikaans. This means that when writing books, authors mostly made use of borrowed words, discarding indigenous Ndebele words. The underlying cause can be the fact that the Ndebele speakers have been in contact with English-, Afrikaans- and Sepedi-speaking peoples. However, from a lexicographic point of view, this theory puts lexicographers in a dilemma, especially when they have to enter and define words as they are used colloquially in everyday speech.

Since 1999, isiNdebele is one of the African languages that have been developed through the dictionary-making process. The isiNdebele lexical stock was selected for and defined in the *IsiNdebele-English Dictionary*. During this project, it has been observed that the traditional approach where the lexicographer operates in a prescriptive capacity using his discretion with regard to the words for inclusion in or exclusion from this dictionary cannot be easily overlooked, especially in this era when the descriptive approach is highly favoured. Traditionally, the lexicographer was seen as judge or guardian of the language. The lexicographer therefore was somebody who chose acceptable words that

should be included in and unacceptable ones that should be excluded from the dictionary with the sole aim of using the dictionary to correct and purify the language.

Metalexicographers such as Al-Kasimi (1977: 84), Prinsloo and De Schryver (2002: 73), Hartmann (1983: 20) and others contend and reiterate that the responsibility of a dictionary is to record the language and not to determine its style. Challenges to the treatment of borrowed words when compiling African language dictionaries have been dealt with in Nong, De Schryver and Prinsloo (2002) for Sepedi and Drame (2001) for isiXhosa and they are all in agreement that the challenges faced with borrowed words lies in language attitude which vary from one language to the other. Ncube (2005) researched the attitude of target users regarding the acceptance and non-acceptance of borrowed words in Zimbabwean Ndebele. His finding was that indeed language attitude is at play because people regard borrowing as a process of language corruption.

In this article, therefore, the lemmatization of loan words in the *isiNdebele-English Dictionary* and the selection of words for lemmatization are discussed. The dictionary under discussion uses the stem-based approach which, De Schryver (2010: 162) refers to it as the 'one-size-fits-all approach' and it has been used by many African languages when compiling dictionaries. In conclusion, the article attempts to determine whether borrowing is successful in the growth and development of isiNdebele, and whether the lexicographer should be the final judge on which borrowed terms to include in the dictionary, or whether language users should be the ones to decide for the lexicographer.

Scholars sometimes use the linguistic terms *borrowing*, *loaning* and *adoption* interchangeably in studies that focus on language contact or language mix, (Mahlangu 2007: 1). Gleason (1956: 397) defines *borrowing* as "the copying of a linguistic item from speakers of another speech form", while Thomason and Kaufman (1988) define *adoption* as "the incorporation of foreign features into a group native language by speakers of that language".

Nkondo (1987), Jafta (1987), Mojela (1991), Madiba (1989) and Van Huysteen (2003) argue that no language is lexically self-sufficient, because no perfectly homogeneous language group exists. Language is like a human being, it is never static, it grows every day and it changes as it grows. Given such a co-existence and also a situation in which the culture of the speakers of one language is regarded as socially and technologically dominant, the major flow of linguistic items will be primarily from the language of the dominant cultural group to that of the dominated speech community.

No fewer than four distinct languages have infiltrated isiNdebele, increasing the lexical stock of the language through language contact and borrowing. The languages affecting isiNdebele are Setswana, Sepedi/Se Sotho sa Leboa, Afrikaans and English. The spread of the Ndebele-speaking people all over white farms in the Highveld region around Middelburg, Belfast, Lydenburg, Standerton, Hendrina and Leandra resulted in most of their borrowed lexical items being drawn from Afrikaans rather than from English (Skhosana 1998: 118).

3. Presenting information in the *IsiNdebele–English Dictionary*

3.1 Entries

The *IsiNdebele–English Dictionary* is a bilingual, bi-directional dictionary, which is divided into two sections. The first section comprises isiNdebele headwords with each headword having an English equivalent, followed by the prefix, abbreviation of the part of speech and the definition(s). The second section comprises of English headwords with each headword having the isiNdebele equivalent, followed by an abbreviated part of speech and the definition(s). Two illustrations of entries from the isiNdebele–English and English–isiNdebele section of the dictionary are given respectively under (1) and (2) below:

- (1) *-dzila (i-iin-)* bz. brass that is worn around the neck/legs by Ndebele women
-gubelo (i-ama-) bz. meat that is roasted after a sheep/goat/cow has been slaughtered
- (2) *-brisket (adj.)* inyama yesifuba yenyamazana
-conclusion (n.) isiphetho, isiqu nto

In the above examples, lemmas are in isiNdebele and English and each lemma is followed by the equivalent or definition.

3.2 Lemmatization of loan nouns in the *IsiNdebele Dictionary*

In the *IsiNdebele–English Dictionary* all words are lemmatized under their stems. This approach was accepted by the Board of Directors of the isiHlathululi-mezwi sesinDebele in 2001. The reason why this lemmatization approach was considered to be the appropriate one is because it shows word and lexical relationships and prevents repetition. While implementing this approach, several observations have been made, mainly with regard to loan words. Most common loan words in isiNdebele are channelled to the Class 9 nasal class which is a singular class of Class 10. This is a similar tendency or strategy applied in Sotho languages where most loan words channelled to Class 9 take their plural form in Class 10. Compare the Se sotho sa Leboa in (3) and isiNdebele examples (4) (a-c) below:

- | | |
|------------------------|---------------------------|
| (3) Class 9 | Class 10 |
| <i>Bêibele</i> 'Bible' | <i>diBêibele</i> 'Bibles' |
| <i>kôpi</i> 'cup' | <i>dikôpi</i> 'cups' |
| <i>tanka</i> 'tank' | <i>ditanka</i> 'tanks' |
| <i>lôri</i> 'lorry' | <i>dilôri</i> 'lorries' |
| <i>têntê</i> 'tent' | <i>ditêntê</i> 'tents' |
| <i>rôkô</i> 'dress' | <i>dirôkô</i> 'dresses' |
| <i>thai</i> 'tie' | <i>dithai</i> 'ties' |
| <i>tôronkô</i> 'jail' | <i>ditôronkô</i> 'jails' |

(4)(a)	Class 9 <i>iBhayibheli</i> 'Bible' <i>ibhigiri</i> 'mug' <i>ikopi</i> 'cup' <i>itanka</i> 'tank' <i>iraba</i> 'rubber' <i>itende</i> 'tent'	Class 10/Class 6 <i>iimBhayibheli / amaBhayibheli</i> 'Bibles' <i>iimbhigiri / amabhigiri</i> 'mugs' <i>iinkopi / amakopi</i> 'cups' <i>iintanka / amatanka</i> 'tanks' <i>iinraba / amaraba</i> 'rubbers' <i>iintende / amatende</i> 'tents'
(4)(b)	Class 9 <i>ikotini</i> 'cotton' <i>idrada</i> 'wire' <i>imbhaji</i> 'jacket' <i>imodere</i> 'car'	Class 10 <i>iinkotini</i> 'cottons' <i>iindrada</i> 'wires' <i>iimbhaji</i> 'jackets' <i>iimodere</i> 'cars'
(4)(c)	Class 9 <i>i-eseyi</i> 'essay' <i>i-albhamu</i> 'album' <i>i-adresi</i> 'address' <i>ijarada</i> 'yard'	Class 6 <i>ama-eseyi</i> 'essays' <i>ama-albhamu</i> 'albums' <i>ama-adresi</i> 'addresses' <i>amajarada</i> 'yards'

Lexicographers therefore find themselves faced with a problem regarding the microstructural elements of the lexical item to be lemmatized, especially with the type of variety. The popular stem-based approach of the *IsiNdebele–English Dictionary* requires that the lexical entries of grammatical information such as prefixes and word categories be entered and bracketed immediately after the head word. Compare in this regard the following isiNdebele examples of Class 9 nouns in (5) below.

- (5) *-kosi* (i-/iim-) (bz): 'chief, traditional leader'
-nja (i-/izi-) (bz): 'dog'
-fene (i-/iim-) (bz): 'baboon'
-pilo (i-/iim-) (bz): 'life'

From the examples in (5) above, it can be seen that the morphological information in the case of nouns, i.e. the singular and plural prefixes immediately succeed the headword and, as already intimated, these commonly comprise the morphological information.

Some loan nouns, especially those that take their plural forms in both Classes 10 and 6, pose the problem of (a) which of the two plural variant forms to enter as morphological information, or (b) whether to include them both. Compare the following isiNdebele loan nouns where in (4)(a) there are those that split their plurality into Classes 10 and 6, while examples (4)(b) and (4)(c) take only one plural form.

The grammatical rules state that "it is only the aspirated Class 5 nouns with a double plural form (i.e. nouns that can form their plural by substituting

the singular prefix with either the plural prefix *ama-* or *iim-*) that lose their aspiration when transposed to Class 10 while they retain their aspiration when they appear in Class 6". From the given examples, it can be observed that there are inconsistencies in the lemmas with these double plural forms.

- (a) In example (4)(a), the Class 9 nouns take their plural forms in Class 6 and 10. On the contrary, in examples (4)(b) and (4)(c), the Class 9 nouns take their plural forms either in Class 10 or Class 6.
- (b) Despite the inconsistencies highlighted, it is found that some of the borrowed words falling in these categories have double plurality in the same class, but their meanings are different. The following are examples of such nouns in Class 9/6
 - (i) *iraba* (rubber) > *amaraba* (rubbers)
iraba (eraser) > *iinraba* (erasers)
 - (ii) *itende* (tent) > *amatende* (tents)
itende (room) > *iintende* (rooms) (room especially for the bride made up of grass mats)
 - (iii) *ibhigiri* (mug) > *amabhigiri* (mugs)
ibhigiri (trophy/cup) > *iimbhigiri* (trophies/cups)(vessel with two handles given as a prize to the winner(s) of a competition)

Although isiNdebele speakers frequently use these nouns as they appear in the examples given in (6) below, the solution would of course be to disregard one of the plural prefixes in the section for morphological information, meaning that either Class 10 or Class 6 is given as plural form. This would imply that these nouns will be entered as shown in (6) below.

(6) Class 5/10		Class 5/6
- <i>Bhayibheli</i> (<i>i-/iim-</i>) (bz)	or	- <i>Bhayibheli</i> (<i>i-/ama-</i>) (bz) 'Bible'
- <i>kopi</i> (<i>i-/iim-</i>) (bz)	or	- <i>kopi</i> (<i>i-/ama-</i>) (bz) 'cup'
- <i>tanka</i> (<i>i-/iim-</i>) (bz)	or	- <i>tanka</i> (<i>i-/ama-</i>) (bz) 'tank'
- <i>raba</i> (<i>i-/iim-</i>) (bz)	or	- <i>raba</i> (<i>i-/ama-</i>) (bz) 'rubber'
- <i>tende</i> (<i>i-/iim-</i>) (bz)	or	- <i>tende</i> (<i>i-/ama-</i>) (bz) 'tent'

In doing this, the lexicographer will have become prescriptive, which modern metalexigraphers regard as the traditional approach. According to Al-Kasimi (1977: 84), Hartmann (1983: 20), and others, 'a responsible and accountable dictionary' has to record the language as it is written and spoken. This implies that the lexicographer should record the speaker's language as it is used, rather than focusing on language correction. If a descriptive approach is adopted, the examples of entries shown in (6) above would have several possible plural forms given in the section for morphological information, rather than two as is the case with most noun entries. Compare the examples in (7) below:

- (7) **Class 5/6/10**
-*Bhayibheli* (*i-/iim-/ama-*) (bz) 'Bible'
-*kopi* (*i-/iin-/ama-*) (bz) 'cup'
-*tanka* (*i-/iin-/ama-*) (bz) 'tank'
-*raba* (*i-/iin-/ama-*) (bz) 'rubber'
-*tende* (*i-/iin-/ama-*) (bz) 'tent'

Unfortunately, during the compilation of the *IsiNdebele–English Dictionary*, lexicographers did not work consistently since for loan words such as those shown in (8) below, only one plural prefix was given. In other words, the dictionary, *inter alia*, prescribes the usage, ignoring influences that isiNdebele might have undergone.

- (8) -*tafula* (*i-/iin-*) bz. 'table'
-*tamati* (*i-/iin-*) bz. 'tomato'
-*teksi* (*i-/ama-*) bz. 'taxi'
-*tende* (*i-/ama-*) bz. 'tent'
-*bhayibheli* (*i-/ama-*) 'Bible'
-*bhayisikili* (*i-/ama-*) 'bicycle'

The lemmas in example (6) above should be entered in the same way as those in example (7) to avoid the inconsistencies shown in the lemmas in example (8).

4. Loan words as 'substitutes' for traditional vocabulary

Rasman (1977: 5) claims that many people reject adoptives, considering them to be an impure form of language. Thus they are referred to as 'borrowing' or 'loaning' which could imply that such words do not belong or only temporarily belong in a specific language.

Borrowing or loaning only occurs when a receiving language integrates a foreign item with the aim of closing an information gap in its vocabulary. Otherwise, there is no need for borrowing or loaning. As soon as a receiving language replaces its own items with new foreign items, it illustrates what is termed a language shift if not a language change. This is exactly the case in isiNdebele where numerous lexical items have infiltrated its lexical stock at the expense of vocabulary that it already owns. Commonly only nouns and verbs are vulnerable to borrowing in most African languages, but isiNdebele has gone beyond this normal boundary and has adopted other word categories such as conjunctions, adjectives and adverbs, especially from Afrikaans (Mahlangu 2007: 109). In the recent crises and accidents of Putco buses transporting people between Mpumalanga and Pretoria along the Moloto road, iKwekwezi FM captured a voice of one commuter who expressed her feelings regarding the conditions of the buses for the Yiza Nendlebe 'Come and Listen' programme. In this oral snippet the commuter enumerates three issues regarding Putco buses, fiercely stating:

- (9) *Die eerste ding*, iimbhesi zakwaPutco ziyagcwala over.
'Firstly, Putco buses are always overloaded.'
Die tweede ding, zimaratha.
'Secondly, they are all wrecked.'
Die derde ding, siyaladelwa emsebenzini.
'Thirdly, we come late at work.'

IsiNdebele equivalent numerals such as *kokuthoma* 'firstly', *kwesibili* 'secondly' and *kwesithathu* 'thirdly' do indeed exist and are known to the commuter, but apparently loan equivalents are preferred to the originals. Compare the following similar additional examples in this regard.

- (10) *-rhanorho* (<Afr. *genoeg*) 'enough'/'sufficient' (instead of *-anele*)
-stararha (<Afr. *stadig*) 'slowly' (instead of *kabuthaka*)
-fenarha (<Afr. *vinnig*) 'quick' (instead of *msinya*)
somara (<Afr. *sommer*) 'just' (instead of *nje*)

It has therefore been observed during the compilation of the bilingual dictionary in question, that speakers discard their own lexical items and substitute them with loan items. This has been one of the challenges that isiNdebele lexicographers had to face, finding themselves between the two views of the dictionary making practice, i.e. prescriptive versus descriptive. Descriptive is concerned with the empirical basis whereas prescriptive is concerned with the genuine purpose of a dictionary. When one looks at the descriptive approach it becomes apparent that it is not viable as a single approach for text production if more than one variant prevails. Prescription is also not entirely a viable approach because it influences future text production activities and give the user one prescribed form to use. However, Bergenholtz (2003), Bergenholtz and Gouws (2010) have now found a solution to this predicament. According to these scholars, a proscriptive approach draws the best of both prescription and description and includes features of both approaches. It is an approach that is used in modern-day dictionaries. However, the focus of this article is not on the prescription, descriptive and proscriptive approaches but they will also be highlighted.

For isiNdebele bilingual dictionary a proscriptive approach is the best approach because it gives a lexicographer the opportunity to include different variants, but also to indicate which one is recommended. A given word can combine with different words in different forms in order to express the same meaning. In isiNdebele the noun *-bhigiri* 'cups' can use both *iim-* and *ama-* prefixes and the noun will be *iimbhigiri* or *amabhigiri* and both express the same meaning.

The foreign lexical stocks such as those in example (6) enjoy the highest frequency of usage amongst isiNdebele speakers. From a descriptive point of view, these words must be lemmatized. The problem is that the so-called National Language Body for isiNdebele (or former old isiNdebele Language

Board) plays a prescriptive role while the lexicographer's purpose has to follow a proscriptive approach.

Under these circumstances, lexicographers certainly find themselves in a dilemma since none of the lexical items such as those in example (10) are found in the electronic corpus of isiNdebele. This corpus is based mainly on written material, which contains formal standardized isiNdebele only, and not much in the line of informal, oral recordings. The question is: Should the foreign items such as those in (10) be lemmatized in the dictionary, replacing original isiNdebele items, as speakers suggest, or not? Should the lexicographers concur with Mdee (1999: 129) when stating that: "A dictionary shall command authority over its users if it convinces them that it is adhering to the standard language. Otherwise it will lose credibility as an authoritative reference of the standard language."

These are some of the questions the isiHlathululi-mezwi sesiNdebele Dictionary Unit, had to answer while the *Bilingual Scholar's Dictionary* was being compiled. If *mará* 'but', *dereveyi* 'while', etc. are incorporated in the dictionary, does it imply that isiNdebele is enriching its vocabulary by borrowing, or indirectly and gradually drifting towards Afrikaans? The lexicographer, in such instances, is bound to be prescriptive, since in spoken language hardly a sentence is uttered in the isiNdebele speech community without an Afrikaans item being used. Compare in this regard the following examples where foreign items frequently replace isiNdebele equivalents.

(11) IsiNdebele	Afrikaans
(Less preferred)	(Highly preferred)
<i>kodwana</i>	<i>mará</i> (<Afr <i>maar</i>) 'but'
<i>ngesimanga sokuthi /</i> <i>ngesikhathi / nagade</i>	<i>dereveyi</i> (<Afr <i>terwyl</i>) 'because of / meanwhile'
<i>eqinisweni</i>	<i>entlege</i> (< Afr <i>eintlik</i>) 'in fact'
<i>ngamane</i>	<i>gamare</i> (<Afr <i>kan maar</i>) 'rather'
<i>mbala</i>	<i>jeyisi</i> (<Afr <i>juis</i>) 'exactly / definitely'

5. Conclusion

In the foregoing discussion, it has been shown how loan items can pose a challenge to lexicographers, especially where they have not yet decided whether they should be prescriptive, descriptive or proscriptive. The article aims to help lexicographers to decide which trend should be followed when they lemmatize words such as those with double plurals in isiNdebele. The suggested trend is the most useful, user-friendly and space-saving because instead of entering the same lemma twice, the lemma is treated once with the double plurals included, except where the plurals have different meanings. The article has also illustrated that loan items differ from one language to another and also that foreign items sometimes behave differently so much so that lexicographers can easily maintain or distort the language usage. The double pluralism that isiNdebele

exhibits in loan words of Class 6 and 9 shows that each language needs to have its specific treatment that will satisfy its users. The article has illustrated that some lexicographic principles such as descriptiveness or prescriptive can be a challenge in a language where its history of recognition is still in its infancy. IsiNdebele has further shown that borrowing is not necessarily an ideal trend towards language enrichment, but it sometimes leads to a language shift or a language impoverishment. The paper has also revealed that borrowing not only contributes towards language growth, as perceived by many scholars, but to some extent, in some indigenous African languages, language diminution or decay, especially when considering how isiNdebele opted to drop some of its own vocabulary at the expense of borrowed Afrikaans vocabulary.

References

- Aitchison, J. 2001. *Language Change: Progress or Decay?* Third Edition. Cambridge: Cambridge University Press.
- Al-Kasimi, A.M. 1977. *Linguistics and Bilingual Dictionaries*. Leyden: E.J. Brill.
- Bergenholtz, H. 2003. User-oriented Understanding of Descriptive, Proscriptive and Prescriptive Lexicography. *Lexikos* 13: 65-80.
- Bergenholtz, H. and R.H. Gouws. 2010. A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography. *Lexikos* 20: 26-51.
- Department of Education and Training. 1983. *IsiNdebele Terminology and Orthography 1*. Pretoria: Government Printers.
- De Schryver, Gilles-Maurice. 2010. Revolutionizing Bantu Lexicography — A Case Study. *Lexikos* 20: 161-201.
- Drame, A. 2001. Foreign Words as Problems in Standardisation/Lxicography: English and Afrikaans Loan-words in isiXhosa. *Lexikos* 11: 231-241.
- Gleason, H.A. 1956. *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart & Winston.
- Hartmann, R.R.K. 1983. *Lexicography: Principles and Practice*. London/New York: Academic Press.
- Iziko lesiHlathululi-mezwi sesiNdebele. 2006. *IsiNdebele-English/English-IsiNdebele IsiHlathululi-mezwi/Bilingual and Explanatory Dictionary*. Cape Town: Phumelela.
- Jafta, N. 1987. The Development of Terminology in Xhosa — A Case Study. *Logos* 7(2): 127-138
- Kamwangamalu, N.M. 1997. Language Frontiers, Language Standardization, and Mother Tongue Education: The Zaire-Zambia Border Area with Reference to the Bemba Cluster. *South African Journal of African Languages* 17(3): 88-94.
- Madiba, M.R. 1989. *A Preliminary Survey of Adoptives in Venda*. Unpublished mini-dissertation for B.A. Honours. Pretoria: University of South Africa.
- Mahlangu, K.S. 2007. *Adoption of Loan Words in isiNdebele*. Unpublished M.A. dissertation. Pretoria: University of Pretoria.
- Mdee, J.S. 1999. Dictionaries and the Standardization of Spelling in Swahili. *Lexikos* 9: 119-134.
- Mojela, V.M. 1991. *Semantic Changes Accompanying Loan-words in the Northern Sotho Lexicon*. Unpublished M.A. Dissertation. Pretoria: Vista University.
- Mojela, V.M. 2010. Borrowing and Loan Words: The Lemmatization of Newly Acquired Lexical Items in Sesotho sa Leboa. *Lexikos* 20: 700-707.

- Ncube, C.** 2005. Language Development or Language Corruption? The Case of Loan-words in *Isichazamazwi SesiNdebele*. *Lexikos* 15: 294-305.
- Ndebele Language Board.** 1995. *Amatjhada/Imidumo Yelimi lesiNdebele nemiThetho yokuTlolwa Kwaso Ndebele Sound System and Orthography*. Pretoria: Via Afrika.
- Nkondo, C.P.N.** 1987. Problems of Terminology in African Languages with Special Reference to Xitsonga. *Logos* 7(2): 69-78.
- Nong, Salmira, Gilles-Maurice de Schryver and D.J. Prinsloo.** 2002. Loan Words versus Indigenous Words in Northern Sotho — A Lexicographic Perspective. *Lexikos* 12: 1-20.
- PanSALB.** 2005. *isiNdebele Revised Orthography and Spelling Rules*. Pretoria: PanSALB.
- Prinsloo, D.J. and Gilles-Maurice de Schryver.** 2002. The Use of Slashes as a Lexicographic Device, with Special Reference to the African Languages. *South African Journal of African Languages* 22(1): 70-91.
- Rasman, E.** 1977. *A Preliminary Survey of Adopting in Southern Sotho*. Johannesburg: Wits University Press.
- Shabangu, T.M. and J.J. Swanepoel.** 1989. *Isihlathululi-mezwi. An English–South Ndebele Dictionary*. Cape Town: Maskew Miller Longman.
- Skhosana, P.B.** 1998. *Foreign Interferences in the Sound, Grammatical and Lexical System of Southern Ndebele*. Unpublished M.A. dissertation. Pretoria: University of Pretoria.
- Skhosana, P.B.** 1999. Dictionary Unit for isiNdebele. *Lexikos* 9: 222-224.
- Southern Ndebele Language Board.** 1982. The Southern Ndebele Orthography. *Educamus* 28(2): 28-31.
- Southern Ndebele Language Board.** 1982. The Southern Ndebele Spelling Rules. *Educamus* 25(5): 35-36.
- Thomason, S.G. and T. Kaufman.** 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Van Huyssteen, L.** 2003. *A Practical Approach to the Standardisation and Elaboration of Zulu as a Technical Language*. Unpublished doctoral thesis. Pretoria: Unisa Press.

Example Sentences in Bilingual Specialised Dictionaries Assisting Communication in a Foreign Language

Sandro Nielsen, *Centre for Lexicography, Aarhus University, Denmark and
International Centre for Lexicography, University of Valladolid, Spain*
(sn@bcom.au.dk)

Abstract: Practitioners, researchers and translators are persons who often have to communicate about domain-specific issues in a foreign language and may consult bilingual specialised dictionaries for help. However, many specialised dictionaries focus on terms even though studies of the foreign language text production process reveal that this also includes grammar, language conventions, genre conventions and style. Specialists can be expected to know conventions and style in their own source language culture but cannot be expected to know how these are realised in a foreign language. Bilingual specialised dictionaries can help users if they contain domain-specific example sentences illustrating how source language convention and style can be transposed to a foreign language. This means that bilingual specialised dictionaries should not merely help users translate terms but be lexicographical tools designed to assist in foreign language communication so that they provide help to practitioners, researchers and translators who produce specialised texts in a foreign language using proper conventions and style.

Keywords: SPECIALISED LEXICOGRAPHY, ONLINE DICTIONARIES, PRINTED DICTIONARIES, TECHNICAL DICTIONARIES, SPECIALISED COMMUNICATION, EXAMPLES, LEXICOGRAPHICAL FUNCTIONS, TEXT PRODUCTION, USER NEEDS, WRITING, TRANSLATION

Opsomming: Voorbeeldsinne in tweetalige vakwoordeboeke help met kommunikasie in 'n vreemde taal. Praktisyns, navorsers en vertalers moet dikwels in 'n vreemde taal oor vakspesifieke sake kommunikeer en raadpleeg daarvoor dan dalk tweetalige vakwoordeboeke. Baie vakwoordeboeke se fokus is op terme alhoewel navorsing oor teksproduksie in 'n vreemde taal daarop wys dat grammatika, taalkonvensies, genrekonvensies en styl ook ter sake is. Van vakspecialiste kan dit verwag word om vertrouwd te wees met taalkonvensies en styl in hulle eie brontaal, maar nie met die weergawe daarvan in 'n vreemde taal nie. Tweetalige vakwoordeboeke kan gebruikers help deur vakspesifieke voorbeeldsinne in te sluit wat wys hoe brontaalkonvensies en styl na die doeltaal oorgedra moet word. Dit beteken dat tweetalige vakwoordeboeke nie net gebruikers moet help om terme te vertaal nie maar hulle moet leksikografiese instrumente wees wat ontwerp is om met vreemdetaalproduksie te help. Daardeur ondersteun hulle praktisyns, navorsers en vertalers wat vaktekste in 'n vreemde taal moet skep om die regte konvensies en styl te benut.

Slutelwoorde: AANLYN WOORDEBOEKE, GEBRUIKERSBEHOEFTE, GEDRUKTE WOORDEBOEKE, GESPECIALISEERDE KOMMUNIKASIE, LEKSIKOGRAFIESE FUNKSIES, SKRYF, TEGNIESE WOORDEBOEKE, TEKSPRODUKSIE, VAKLEKSIKOGRAFIE, VERTALING, VOORBEELDE

1. Introduction

As a result of increasing activities across borders and cultures, more and more people need to communicate in a foreign language in a range of situations related to specific subject fields. Translators often help businesses and specialists communicate in foreign languages. Practitioners from various domains have become more internationally oriented because they work for multinational organisations, have been posted or seconded abroad. Researchers from all types of disciplines study not only their local environments but also look for inspiration in foreign environments to a greater extent than ever before and therefore must be able to understand foreign language source material. In addition, practitioners as well as researchers are expected to act in international settings by publishing articles in journals, by presenting papers at conferences, and by participating in negotiations or discussions. It is imperative that translators and domain specialists communicate correctly in a foreign language about the subject matter concerned and in this light, it is relevant to examine whether, and if so how, dictionaries can provide help in communicative situations.

Specialised dictionaries can contain a great variety of data types that help practitioners, researchers and translators successfully meet challenges in foreign language communication. Equivalent terms are the obvious type of data, and so are examples of language use. Example sentences represent one type of example that have been discussed in the lexicographical literature, but the study of example sentences in specialised dictionaries designed to provide help in communicative situations in a foreign language has so far been relatively limited. This paper discusses how example sentences in bilingual specialised dictionaries can facilitate domain-specific communication in a foreign language. The following discussion first delimits the scope of examples in lexicographical products and defines the concept of specialised dictionaries, then looks at the relationship between dictionaries and the production of written texts in a foreign language, and finally discusses how lexicographers may use example sentences to assist translators, researchers and practitioners in their foreign language communication.

2. Defining examples and specialised dictionaries

One type of lexicographical data that lexicographers can include in their dictionaries with a view to helping practitioners, researchers and translators com-

municate in a foreign language is often referred to as examples. A study of the existing literature reveals that the general term *example* is used in more than one sense. For instance, Svensén (2009: 281) explains that the term "comprises all types of indications consisting of word combinations (phrases, clauses, sentences) that contain a form of the lemma sign, irrespective of whether they are provided with an indication of meaning or not." This description applies, in particular, to monolingual dictionaries, such as dictionaries intended for text production by foreign language users, though bilingual dictionaries with the same function may also contain lexicographical examples. In order to narrow down the discussion that follows, focus will be on sentences, mainly because they may illustrate the use of equivalents in context as well as the translation of lemmas (or headwords) in context relevant for specialised communication in a foreign language. Another reason is that collocations and phrases may exemplify the use of lemmas in monolingual dictionaries, but in bilingual dictionaries "the majority of sub-entries provide one-to-one translation equivalents. They exemplify nothing" (Jacobsen et al. 1991: 2783). Finally, a recent study shows that "The data support the conclusion that phrases are less useful at providing support than sentences" (Hiles 2011: 304).

Lexicographers often divide examples into different types. Zöfgen (1986) proposes a classification based on the properties or characteristics of lemmas in monolingual dictionaries for learners such as, examples showing grammar data, collocations, clauses and phrases. Harras (1989) distinguishes between syntactic data, collocations and sentences illustrating usage in monolingual dictionaries, among others. Furthermore, Bergenholtz (1995: 137-142) discusses examples in specialised dictionaries and distinguishes between collocations, grammatical examples, citations, citation examples and competence examples; overall, he favours the use of example sentences.

One reason for the different understandings of and approaches to examples in dictionaries appears to be the influence of linguistics and lexicology. Jacobsen et al. (1991: 2784), who address examples in bilingual general dictionaries, suggest that examples should not be examined using a functional approach (understood as linguistic functions) but from a formal perspective (understood as components in dictionaries); in their words, an example in a dictionary cannot "be functionally defined in terms of the type of information it provides; it follows from this that it is not (meta)linguistically defined in terms of the linguistic categories (syntax, semantics, morphology) involved" (Jacobsen et al. 1991: 2784). After this brief discussion of examples in lexicography, it is appropriate to provide a working definition of lexicographical example sentences intended to assist translators and specialists communicate in a foreign language:

Example sentences are full sentences that provide data concerning the source language lemma, the target language equivalent or a combination of both enabling specific types of users to communicate successfully in a foreign language in domain-specific situations.

A specialised dictionary, whether printed or online, is a complex unit. Firstly, the dictionary has several surface features, i.e. the printed or online dictionary contains a number of distinct components that constitute the dictionary when users consult it. For example, printed dictionaries contain wordlists as well as other supporting components such as appendices and user guides, while online dictionaries contain components such as search sites, results sites, and various supporting components. Secondly, the dictionary has three underlying features, namely:

- (1) it has been designed to fulfil one or more functions, i.e. provide a specific type of help in specific types of situation to specific types of user;
- (2) it contains data that have been selected because they help to fulfil its function(s);
- (3) it has structures that organise its data into the task of fulfilling its function(s).

Specialised dictionaries can be designed to help users in a range of situations, of which two are the most general types. According to Nielsen (2012: 114), dictionaries with communicative functions provide help in ongoing or planned communicative situations, while dictionaries with cognitive functions provide help to those who want to acquire knowledge about factual or linguistic matters. Specialised dictionaries can therefore be designed to:

- provide help to translate specialised texts
- provide help to produce specialised texts
- provide help to understand specialised texts
- provide help to acquire general or specific knowledge about factual or linguistic matters from one or more subject fields

This means that specialised dictionaries are lexicographical tools that function as utility products which, through their surface and underlying features, provide specific types of help to specific types of user in specific types of usage situation related to one or more domains and the related domain-specific language. In order to make such information tools, it is relevant to study the extra-lexicographical situation when translators, researchers and practitioners produce written texts in a foreign language.

3. Dictionaries and text production in a foreign language

People who need to communicate about domain-specific matters in a foreign language will need dictionaries that can help them solve specific problems. At a very general level this may be summed up as follows: the data contained in the dictionaries (e.g. in example sentences) should enable practitioners, researchers

and translators to write grammatically and idiomatically correct utterances without compromising their factual contents. Bergenholtz and Nielsen (2002: 14) suggest that the general purpose of dictionaries intended for text production is to help users become aware of the elements of good writing within a subject field; this includes aspects such as grammatical, syntactic, stylistic and terminological questions that arise during the writing process. These aspects are important when translating source texts into a foreign language as well as when writing texts directly in the foreign language since the only major step in text production that distinguishes the two types of situation is the analysis of source language texts to be translated. When identifying the steps in producing texts, lexicographers should use a method that is workable in lexicography.

A broad outline of the writing process may be used to determine the choices that lexicographers have to make. Marsen (2003: 2-13) explains that the process of writing texts has three stages: a planning, an execution and a finalisation stage. In most cases, writing manuals provide help in the planning stage so lexicographers should focus on the execution and the proofreading stages. Rude (2002: 15-16) shows that the execution stage involves the writing of a draft text, followed by the revising and editing of the text, and that proofreading extends revising and editing into the finalisation stage. Similarly, Byrne (2006: 17) discusses technical translation and breaks down the translation process into five tasks: interpretation of the source text, writing of the translation, research needed for the first two tasks, checking the draft translation for errors, and making any necessary corrections and adjustments. Specialised dictionaries for text production can thus be described as tools that provide help to draft, revise and edit specialised texts.

Text production occurs at two general levels. The macro-level concerns paragraphs and larger units of text, whereas the micro-level concerns words, collocations, phrases, sentences, and textual conventions. The micro-level is particularly relevant to specialised dictionaries because specialists and translators need help at this level when writing texts in a foreign language. Mossop (2007: 27-33) explains that when they revise and edit, writers make sure that their texts follow generally accepted grammar and spelling rules, that they have used the correct terminology, and that the finished texts are unambiguous. Finally, writers and copyeditors look for errors and correct them as appropriate. The process of specialised translation involves similar aspects: "Technical translation involves detailed knowledge of the source and target cultures, target language conventions, text type and genre conventions, register, style, detailed understanding of the audiences" (Byrne 2006: 7). It follows that specialists and translators producing texts in a foreign language are challenged at the micro-level and lexicographers should therefore focus on this level when making dictionaries designed to provide help to write and translate specialised texts.

When selecting data types, lexicographers should take into consideration the competences of users. By studying the foreign language text production process and by identifying the relevant skills and competences of practitioners,

researchers and translators, lexicographers have a good basis for selecting data that can best satisfy user needs. Several methods for profiling intended dictionary users exist but few focus on specialised dictionaries. Bergeholtz and Nielsen (2006: 286) propose a list of questions that relate to categories of skills and knowledge possessed and for the discussion below, the following overall competences are assumed:

- Practitioners, researchers and translators master their native language at a high level
- Practitioners and researchers master the foreign language at a low or medium level
- Translators master the foreign language at a high level
- The experience of practitioners and researchers in producing domain-specific texts in the foreign language is at a low level
- The experience of translators in producing domain-specific texts in the foreign language is at a medium or high level
- The general cultural and factual knowledge of practitioners, researchers and translators is at a high level
- Practitioners and researchers master their subject field in their native culture at a high level
- Translators master a subject field in their native culture at a medium level
- Practitioners, researchers and translators master their culture-dependent subject field in the foreign culture at a low level (or not at all)
- Practitioners, researchers and translators master their culture-independent subject field in the foreign culture at a medium or high level
- Practitioners and researchers master the native language of their subject field at a high level
- Practitioners and researchers master the language of their subject field in the foreign culture at a low or medium level
- Translators master the two languages of the native and foreign subject fields at a medium or high level

A list like the above enables lexicographers to respond to the general needs of translators, researchers and practitioners who have to communicate in a foreign language. It would appear from the above list that, in communicative situations, some will think in their native language and translate into the foreign language, others will think in their native language and translate into the foreign language as well as think in the foreign language and write directly in the foreign language, and yet others will think in the foreign language and

write directly in the foreign language. Lexicographers should take these factors into consideration when deciding on the type and form of example sentences that are intended to help users communicate about their subject field in a foreign language.

4. Example sentences supporting text production

When communicating in a foreign language, three general topics come into play: terms, language structures, and genre conventions. These topics are particularly important when the native language of translators and specialists allows the use of terms, language structures and genre conventions that differ from the functionally equivalent terms, structures and conventions in the foreign language. Furthermore, as example sentences are dictionary components, lexicographers should be able to use such sentences to show or translate all three as they are found at the micro-level of text production. This is in line with the comment made by Jacobsen et al. (1991: 2787) in respect of general dictionaries, "Examples, if judiciously selected, can demonstrate collocational, stylistic, syntactic, morphological, and cultural features of words and phrases." If this statement is compared to the description by Byrne (2006) quoted above (that technical translation requires knowledge of the two cultures involved, target-language conventions, genre conventions, register, and style), lexicographers should consider using example sentences to show how specialised communication can be performed in a foreign language.

Style, i.e. the characteristic and correct way of speaking and writing in a specific domain, may be regarded as an overarching feature of specialised text production. One of the misconceptions about technical and other specialised texts is that style does not matter, since the factual content overshadows everything else; however, according to Byrne (2006), this misconception is completely unfounded:

Perhaps the problem stems from differing opinions of the nature of style and the popular belief that it relates exclusively to literature. If we look at style from a literary point of view, then it does not have any place in technical translation. But if we regard style as the way we write things, the words we choose and the way we construct sentences, the style is equally, if not more, important in technical translation than in other areas because it is there for a reason, not simply for artistic or entertainment reasons. (Byrne 2006: 4)

As a result, lexicographers should allow room for style in dictionaries intended to help specialists and translators produce domain-specific texts in a foreign language. Example sentences may represent or exemplify the grammatical properties where irregularities are concerned, as experts, researchers and translators can only be expected to know regular grammatical rules of the foreign language. One example sentence will usually suffice but in some cases, lexicographers should consider whether two or more sentences would be appro-

priate. Consider the following two sentences exemplifying the use of the British legal term *counsel* (i.e. lawyer):

- The judge asked counsel for the defence to explain. (*Longman Dictionary of Contemporary English*)
- The plaintiff appeared in court with his solicitor and two counsel. (*English Law Dictionary*)

The first example sentence demonstrates that *counsel* is used correctly without an article when it refers to a lawyer, even when followed by a post-modifying prepositional phrase. The second example sentence shows that *counsel* inflects irregularly for the plural, as no inflectional affix is added to the stem. It would be difficult to find (or construct) a typical or natural sentence showing both grammatical irregularities and at the same time be easy to understand. However, the danger of including too much data in example sentences is that users may overlook or misinterpret the data because they do not know what to look for.

Lexicographers can also use example sentences to show incorrect use of terms. *A Dictionary of Modern Legal Usage*, intended to help lawyers produce better legal writings, often provides examples of incorrect use to make lawyers aware that specific terms are often used wrongly in practice. The article treating the term *counsel* contains the following authentic example sentence with the lexicographers comment in square brackets:

- [F]our lawyers were named Nov. 25 to serve as *counsels* [read *counsel*] for the transition. (*A Dictionary of Modern Legal Usage*)

Even though they are from monolingual dictionaries, the above three example sentences exemplify the use of specialised language. Such illustrations are useful to both native and non-native speakers of English; the main reason is that the specialised language of domains use words, structures and conventions that differ from those in everyday language and that specialist style allows the use of structures and conventions that are seldom, if ever, used in general language. In bilingual dictionaries, the three English example sentences above will illustrate the use of the equivalent (in this case *counsel*) and thus be addressed to the equivalent, whereas the sentences would be addressed to the lemma in monolingual dictionaries.

An issue that has been discussed in the literature is whether example sentences in bilingual dictionaries should be in the source language (SL) as well as in the target language (TL). Adamska-Sałaciak (2006: 493-494) "believe[s] that there is no need to translate examples" but modifies this by saying that "As far as systematic translation of all examples is concerned, I can see some justification for the practice in dictionaries targeted at beginners." These comments concern general dictionaries, but if domain-specific communication contains terms, structures and conventions that are seldom used in general language, lexicographers may have additional justification for translating example sen-

tences in specialised dictionaries. The existing literature also contains different opinions and positions as to the translation of examples in general dictionaries, for instance:

In bilingual lexicography it is not sufficient to refer to the "SL example" or the "TL example" for a given lemma. The SL and TL halves of the example are closely linked functionally, and must be considered to be a unit formally. If the primary focus of the bilingual dictionary is translation, then the primary function of the bilingual dictionary example is to exemplify translation. A definition of the form of an example must logically include its SL and TL halves. (Toope 1996: 18)

An example may clarify the position. The following English sentence may be a target-language example sentence in an imaginary German–English technical dictionary:

- Figure 1 contains the characteristics of diodes of equal power rating. (Hann 1992: 190)

This sentence illustrates the use of an equivalent in context, but tells users nothing about the relationship between this sentence and any source language utterance. Only implicitly does the English sentence show how a German utterance with a similar factual message (provided users can come up with one) may be translated — the SL half of the example sentence is missing. Nevertheless, the sentence does show one instance of target-language usage. If it contains example sentences in SL as well as their translations in TL, the German–English technical dictionary may provide the following example sentence:

- In Abbildung 1 ist das Kennlinienfeld leistungsgleicher Dioden angegeben.
- Figure 1 contains the characteristics of diodes of equal power rating. (Hann 1992: 190)

Now the two halves of the example sentence are present and it becomes clear that the sentence shows stylistic differences in technical discourse between the two languages, in that the German sentence begins with a prepositional phrase and is phrased in the passive voice. In contrast, the English translation shows that target language style conventions prefer the use of an active construction with no prepositional phrase. This difference is made explicit only when the SL half and the TL half of the example sentence are presented. Similarly, an imaginary Danish–English technical dictionary may contain the following example sentence:

- En indstillelig momentkobling slår spindlen fra før tapbrud.
- Adjustable torque control stops spindle to prevent tap breakage. (Thürmer 1986: 135)

This example shows an English equivalent in context and explicitly illustrates a stylistic difference between the two languages. The Danish half contains an indefinite article ("En indstillelig") and a definite article ("spindlen"). In contrast, the English half shows that English conventions prefer the non-use of articles when correctly expressing the meaning of the Danish utterance. The example sentence also shows another difference: Danish apply general-language conventions allowing the use of articles, whereas English apply special-language conventions that do not include articles in this particular context. Based on their assumed skills and competences, translators, researchers and practitioners are unlikely to be aware of such differences, something which justifies the inclusion of bilingual example sentences.

There is no doubt that lexicographers can use example sentences to illustrate style conventions in a foreign language, but users may not immediately notice that example sentences show such conventions. So the challenge is for lexicographers to present style conventions in a way so that users will not have to guess that example sentences provide data that help to produce foreign language utterances complying with domain-specific style conventions. If the above imaginary bilingual dictionaries are printed lexicographical tools, lexicographers may place the example sentences in specific boxes with informative headlines; see Example 1.

Style convention: passive and active sentences
In Abbildung 1 ist das Kennlinienfeld leistungsgleicher Dioden angegeben.
Figure 1 contains the characteristics of diodes of equal power rating.

Example 1: Example sentence with informative headline in German–English technical dictionary

The imaginary dictionary may also be an online lexicographical tool and lexicographers may present the example sentence in the same way as illustrated in Example 1. However, lexicographers can use other features to explicitly inform users of the illustrative function; for instance, the example sentence may be placed in a box with an icon, as shown in Example 2.

In Abbildung 1 ist das Kennlinienfeld leistungsgleicher Dioden angegeben.
Figure 1 contains the characteristics of diodes of equal power rating.



Example 2: Example sentence accompanied by icon and possible hover box

Icons may be combined with hover boxes, also called mouse-overs or mouse hovers, so that users place the pointer over the icon and a box containing, e.g.

the text "Style conventions: passive and active sentences", appears without the need to click.

Lexicographers may also present example sentences in printed and online dictionaries using boldface, italics and different colours for specific style conventions in the SL and/or TL half. The use of colours should be accompanied by colour legends, e.g. red = grammar; yellow = culture; blue = textual convention. Alternatively, lexicographers can show the meaning of colours through hover boxes so that users place the pointer over a coloured text part and a box automatically appears with an informative text. In online dictionaries, lexicographers may further consider to let users press the coloured text parts on a touchscreen instead of using pointers to make informative hover boxes appear.

A related but slightly different approach lexicographers can use in both printed and online dictionaries is to present entire SL sentences but only partial translations, as illustrated in Example 3.

appropriate adj. 1. geeignet, entsprechend, sinnvoll, passend: [.....] To assist in selecting **the most appropriate procedure**, let us consider the advantages and disadvantages of each alternative (*passendste Vorgehensweise*). [.....] **It is appropriate to** specify that the control system be able to handle 1 000 analog inputs and outputs (*es ist zweckmäßig/sinnvoll vorzuschreiben, dass ...*).

Example 3: Partially translated example sentences from *Phraseological Dictionary English–German*

Example 3 shows how text parts of SL sentences can be highlighted in boldface and how these passages can be translated and presented in italics. The first example sentence illustrates the different ways to compare adjectives in English and German, while the second sentence shows that an English lemma has two possible translations in German, of which one is given as an alternative to the equivalents presented in the head of the article. The advantage of highlighting passages in SL and TL halves is that lexicographers draw attention to the stylistic, linguistic or genre-related points they want to convey to users (Prinsloo and Gouws 2000: 148), but a disadvantage is that users may be distracted by the linguistic and factual complexity of relatively long sentences.

Writers of specialised texts know that the use of textual or genre conventions vary considerably. Mayoral Asensio (2007: 52) describes a genre as a class of texts perceived as such by readers of the texts who recognise conventions regarding textual structure and other linguistic elements used in similar communicative situations. As indicated in Section 3 above, lexicographers should consider those conventions that appear at the micro-level since they are likely to differ from one genre and domain to another and since different cultures and languages have their own ways in which to realise the conventions. Practitio-

ners, researchers and translators therefore need dictionaries that help them produce texts that conform to the conventions in the target language, because "Failing to comply with target language text conventions can undermine the credibility of the text, the author and the information in the text" (Byrne 2006: 4).

Specialists and translators are generally expected to be familiar with genre-specific conventions in their own language within one or few subject fields but only few genre conventions in a foreign language. Nielsen (2000) discusses genre conventions used in statutes and suggests that since culture dictates convention, good bilingual translation dictionaries should help users select the proper foreign language conventions. Law is a culture-dependent subject field in that the legal structure and the legal language in one culture differs from the structure and language in another culture; hence law is one of the fields in which those who communicate in a foreign language need help. Example 4 contains the same sentence in seven languages and is from a judgment passed by the Court of Justice of the European Union; it shows the use of textual conventions when referring to European Union directives.

Danish	Direktivets artikel 14, stk. 1, litra c), har følgende ordlyd:
English	Article 14(1)(c) of that directive is worded as follows:
French	L'article 14, paragraphe 1, sous c), de ladite directive est libellé comme suit:
German	Art. 14 Abs. 1 Buchst. c der Richtlinie 2006/54 bestimmt:
Spanish	El artículo 14, apartado 1, letra c), de esa Directiva tiene la siguiente redacción:
Swedish	Enligt artikel 14.1 c i samma direktiv gäller följande:
Dutch	Artikel 14, lid 1, sub c , van die richtlijn is als volgt geformuleerd:

Example 4: The use of textual conventions in different languages in a regulatory genre

The textual conventions in Example 4 have been highlighted using boldface and the sentences show that languages realise the conventions in different ways. The English and the Swedish conventions differ most from those of the other languages in that they are the shortest passages with only one word (term). The other languages use combinations of words and abbreviations with or without brackets, commas and full stops. Practitioners, researchers and translators cannot generally be expected to know such specific textual conventions in a foreign language. Accordingly, lexicographers should consider including example sentences in their dictionaries that help users write foreign language textual conventions when usage in the source language and the target language differs. An imaginary French–English dictionary of law may contain the example sentence shown in Example 5.

L'article 14, paragraphe 1, sous c), de ladite directive est libellé comme suit:
--

Article 14(1)(c) of that directive is worded as follows:
--

Example 5: Example sentence providing help to translate French textual conventions into English

Example 5 illustrates that example sentences can help specialists and translators select the correct way in which to realise conventions in a foreign language. By highlighting the relevant passages containing the conventions, lexicographers draw the attention of users to the genre-related points they want to convey and users receive explicit help with explicit points.

The lexicographical structure that determines the interplay between data in dictionaries is the distribution structure. If lexicographers decide to place the data on grammar, textual conventions and style inside articles, users will find the data in close proximity to the lemmas. Consequently, users will have direct access to these types of data through the wordlist in printed dictionaries and some online dictionaries or through the search for lemmas or words in examples in online dictionaries. However, the data types have to be included in articles in addition to other data types (e.g. morphological data, definitional data, equivalents, collocations, synonyms and cross-references) leading to possible data overload. Furthermore, lexicographers will have to include identical or near-identical grammar, genre-related and style data in a large number of articles and this may increase the length of articles, which takes up space in printed dictionaries. Spatial constraints are not a problem in online dictionaries, but if articles or other search results take up more space than can be shown on a screen, users may overlook important data.

When lexicographers have placed functionally relevant data in the dictionary or database, the next question is how users will be able to find the data. In printed dictionaries, example sentences will be scattered all over the book and in order to help practitioners, researchers and translators find them lexicographers may include one or more indexes in the front or back matter. This solution is used in *A Dictionary of Modern Legal Usage*, which has a "Classified Guide to Essay Entries" in its front matter divided into 5 sections: style; grammar and usage; legal lexicology and special conventions; word formation, inflection, spelling and pronunciation; punctuation and typography. Each section contains a list of lemmas so that users who look for data on, for instance, style, can see in which articles they can find style data related to the lemmas. If it contains example sentences dealing with grammar, textual conventions and style, a bilingual specialised dictionary may contain an index divided in three sections, or three indexes, listing the articles that include example sentences on grammar, textual conventions and style, respectively.

Lexicographers may opt for similar solutions in online dictionaries, i.e. presenting one or more indexes listing the lemmas to which example sentences are addressed. Users can then click on lemmas and be brought directly to full

articles containing the relevant example sentences. Online technology allows lexicographers to adopt other solutions where, for instance, users click lemmas in indexes and are only presented with e.g. lemmas, definitions and example sentences, i.e. not full articles in the traditional sense, but merely those data types addressed to lemmas that are relevant for understanding example sentences. This option will often allow practitioners, researchers and translators to find example sentences on the screen without having to scroll. This means that, in online dictionaries, it does not matter where in the database example sentences are placed; the important point is for lexicographers to select a data presentation structure that presents example sentences dealing with grammar, textual conventions and style so that they are easy to find. Moreover, online dictionaries may also allow users to search for words and terms directly in all example sentences in the database and then present those sentences (in both SL and TL) that contain the words and terms searched for.

5. Concluding remarks

This paper has discussed how example sentences in specialised dictionaries can facilitate text production in and translation into a foreign language. When practitioners, researchers and translators communicate about domain-specific matters in a foreign language they need help with terms as well as other textual elements at the micro-level of text production. By studying the text production process, lexicographers can identify where the challenges are, and it appears that in addition to terms specialists and translators need help with e.g. cultural differences, language conventions, genre conventions and style. One way in which dictionaries can provide help with these elements of foreign language text production is when lexicographers include example sentences in the source language and their translation in the target language that illustrate and exemplify how practitioners, researchers and translators can communicate properly, i.e. grammatically, idiomatically and stylistically correct. To some extent, lexicographers can adopt the same or similar solutions in both printed and online specialised dictionaries, but online technology offers additional ways in which to give easy access to example sentences. This means that lexicographers should not regard bilingual specialised dictionaries as lists of terms and their translation but as lexicographical information tools that provide help to practitioners, researchers and translators who produce specialised texts in a foreign language with grammar, textual conventions and style.

References

Dictionaries

A Dictionary of Modern Legal Usage = Garner, B.A. 1995. *A Dictionary of Modern Legal Usage*. Second edition. New York/Oxford: Oxford University Press.

- Dictionary of Law* = Collin, P.H. 1993. *English Law Dictionary*. Teddington: Peter Collin Publishing.
- Longman Dictionary of Contemporary English* = Summers, D. (Ed.). 2003. *Longman Dictionary of Contemporary English*. Fourth edition. Harlow: Longman.
- Phraseological Dictionary English–German* = Kraus, R. and P. Baumgartner. 2011. *Phraseological Dictionary English–German. General Vocabulary in Technical and Scientific Texts*. Berlin/ Heidelberg: Springer.

Other literature

- Adamska-Salaciak, A.** 2006. Translation of Dictionary Examples — Notoriously Unreliable? Corino, E. et al. (Eds.). 2006. *Atti del XII Congresso Internazionale di Lessicografia, Torino, Italia, September 6th–9th 2006*: 493-501. Alexandria: Edizioni dell'Orso.
- Bergenholtz, H.** 1995. Linguistic Information. Bergenholtz, H. and S. Tarp (Eds.). 1995. *Manual of Specialised Lexicography. The Preparation of Specialized Dictionaries*: 111-142. Amsterdam/Philadelphia: John Benjamins.
- Bergenholtz, H. and S. Nielsen.** 2002. Terms in the Language of Culture-dependent LSP Dictionaries. *Lexicographica. International Annual for Lexicography* 18: 5-18.
- Byrne, J.** 2006. *Technical Translation: Usability Strategies for Translating Technical Documentation*. Dordrecht: Springer.
- Court of Justice of the European Union.** 2014. *Case C- 167/12, C.D. v S.T.* [online]. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0167&qid=1399112550800> [Accessed 15 April 2014].
- Hann, M.** 1992. *The Key to Technical Translation. Volume 2: Terminology/Lxicography*. Amsterdam/Philadelphia: John Benjamins.
- Harras, G.** 1989. Zu einer Theorie des lexikographischen Beispiels. Hausmann, F.J. et al. (Eds.). 1989. *Dictionaries. An International Encyclopedia of Lexicography. Vol. 1*: 607-614. Berlin: Walter de Gruyter.
- Hiles, L.** 2011. Categorising Example Sentences in Dictionaries for Research Purposes. *Lexikos* 21: 298-304.
- Jacobsen, J., J. Manley and V.H. Pedersen.** 1991. Examples in the Bilingual Dictionary. Hausmann, F.J. et al. (Eds.). 1991. *Dictionaries. An International Encyclopedia of Lexicography. Vol. 3*: 2782-2789. Berlin: Walter de Gruyter.
- Marsen, S.** 2003. *Professional Writing. The Complete Guide for Business, Industry and IT*. Basingstoke: Palgrave Macmillan.
- Mayoral Asensio, R.** 2007. Specialised Translation. A Concept in Need of Revision. *Babel* 53(1): 48-55.
- Mossop, B.** 2007. *Revising and Editing for Translators*. Manchester/Kinderhook: St. Jerome Publishing.
- Nielsen, S.** 2000. Translation Strategies for Culture-Specific Textual Conventions in Bilingual Dictionaries. *Lexicographica. International Annual for Lexicography* 16: 152-168.
- Nielsen, S.** 2012. Online Law Dictionaries: How to Provide Help for EFL Text Production by Law Students. *Fachsprache/International Journal of Specialized Communication* 23(3-4): 112-124.
- Prinsloo, D.J. and R.H. Gouws.** 2000. The Use of Examples in Polyfunctional Dictionaries. *Lexikos* 10: 138-156.
- Rude, C.D.** 2002. *Technical Editing*. New York: Longman.

- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Thürmer.** 1986. *Dansk gevindskæreværktøj siden 1898. Danish Thread Cutting Tools since 1898*. Frederiksberg: Thürmer & Co.
- Toope, M.** 1996. *Examples in the Bilingual Dictionary*. Unpublished M.A. Thesis, University of Ottawa, School of Translation and Interpretation, Ottawa. [Online]. Available at: <http://www.ruor.uottawa.ca/en/bitstream/handle/10393/9942/MM15769.PDF?sequence=1> [Accessed 15 April 2014].
- Zöfgen, E.** 1986. Kollokation — Kontextualisierung — (Beleg-)satz. Anmerkungen zur Theorie und Praxis des lexikographischen Beispiels. Barrera-Vidal, A. et al. (Eds.). 1986. *Französische Sprachlehre und bon usage. Festschrift für Hans Wilhelm Klein zum 75. Geburtstag*: 219-238. München: Max Hueber.

The Lemmatisation of Nouns in Tshivenda Dictionaries

Mashudu Nthambeleni (*mashudu.nthambeleni@univen.ac.za*)
and

Nelson Mbulaheni Musehane (*musehanen@univen.ac.za*)
*M.E.R. Mathioha Centre for African Languages, Arts and Culture,
School of Human and Social Sciences, University of Venda,
Thohoyandou, Limpopo, South Africa*

Abstract: Tshivenda is one of the official languages of South Africa, mostly spoken in Limpopo Province at Vhembe District. From 1994, when the Republic of South Africa became a democratic government, speakers of the Tshivenda language spread to all nine provinces of South Africa. Tshivenda has different word categories, like nouns, verbs, adjectives and many more. When words are looked up in dictionaries, one should know what types of categories are given, e.g. their spelling, pronunciation and meaning. If this is not clearly represented one would not be able to use such a dictionary. This paper seeks to investigate how nouns are lemmatised in Tshivenda dictionaries. Attention is given to word and stem lemmatisation. It also looks at the lemmatisation of singular and plural nouns, and the lemmatisation of deverbative and diminutive nouns. This will be accomplished by analysing published Tshivenda dictionaries in circulation.

Keywords: LEMMATISATION, NOUNS, TSHIVENḌA, DEVERBATIVE NOUNS, DIMINUTIVE NOUNS, SINGULAR, PLURAL

Opsomming: Die lemmatisering van substantiewe in Tshivenda woordeboeke. Tshivenda is een van die amptelike tale van Suid-Afrika wat hoofsaaklik in die Vhembedistrik van die Limpopoprovinsie gepraat word. Sedert 1994, toe die Republiek van Suid-Afrika 'n demokratiese regering gekry het, het sprekers van Tshivenda na al nege provinsies van Suid-Afrika versprei. Tshivenda het verskeie woordsoorte, soos substantiewe, werkwoorde, adjektiewe en so meer. Wanneer woorde in woordeboeke nageslaan word, moet 'n mens weet watter datatipes aangebied word, bv. spelling, uitspraak en betekenis. As dit nie duidelik weergegee word nie, sal 'n mens nie in staat wees om so 'n woordeboek te gebruik nie. Hierdie artikel wil ondersoek hoe substantiewe in Tshivenda woordeboeke gelemmatiseer word. Daar word aandag geskenk aan woord- en stamlemmatisering. Daar word ook gekyk na die lemmatisering van enkel- en meervoudige substantiewe, asook die lemmatisering van deverbatiewe en diminutiewe substantiewe. Dit word gedoen deur bestaande Tshivenda woordeboeke te ondersoek.

Slutelwoorde: LEMMATISERING, SUBSTANTIEWE, TSHIVENḌA, DEVERBATIEWE SUBSTANTIEWE, DIMINUTIEWE SUBSTANTIEWE, ENKELVOUD, MEERVOUD

1. Introduction

If dictionaries are not compiled according to the required acceptable standard,

users of such dictionaries would find it difficult to use them. In other words such dictionaries will not be user-friendly. There should be proper planning to determine what information about spelling, pronunciation as well as meaning should be included. This process is termed *lemmatisation*. Hartmann and James (1998: 83) define *lemmatisation* as "the reduction of a paradigm of variant word forms to a canonical form; e.g. the inflected forms of English verbs to the infinitive." They add that lemmatisation is the process by which the compilers of dictionaries establish the canonical forms of headwords either by removing or retaining inflections.

On the other hand, Plisson et al. (2005: 369) define lemmatisation as "the process of finding the normalized form of words" as their arrangement into alphabetical order whereas Sinclair (1991: 173) defines lemmatisation as "the process of gathering word-forms and arranging them into lemma or lemmata".

From these definitions, it can be deduced that lemmatisation is the selection of words or data to be included in a dictionary. According to Gove (1961: Preface 4a), the lemmatisation of words is determined by the degree to which they are most likely to be looked for.

A noun in Tshivenda is *dzina* "name". According to the *Macmillan English Dictionary for Advanced Learners* (2002: 967) a noun is a word or group of words used for referring to a person, thing, place or quantity whereas the *Chambers-Macmillan South African Student's Dictionary* (1996: 639) defines a noun as a word used to refer to a person, thing or quality. By that is meant that it can refer to anything that can occupy a space and have weight as well as some abstract qualities.

Several wordlists and dictionaries have been published in Tshivenda, e.g. those by Marole (1955), Marole and De Gama (1958), Joubert and Rapea (1972), Wentzel and Muloiwa (1982), Hartshorne (1984), Van Warmelo (1989) and that by the Tshivenda National Lexicography Unit, *Tshivenda/English Thalusamaipfi/Dictionary* (henceforth abbreviated as TETD) (2006). This article will analyse the lemmatisation of nouns, particularly in Wentzel and Muloiwa (1982), Van Warmelo (1989) and TETD (2006).

When the current dictionaries of South African languages, including Tshivenda, are investigated, it is found that nouns are lemmatised principally in two ways:

- (a) using the whole word (noun), and
- (b) using the noun stems.

2. Lemmatisation of nouns

Lemmatising the whole word in languages such as Tshivenda, Sesotho sa Leboa and Xitsonga has become tradition. Lemmatising according to the noun stem is found in the Nguni languages, where it originated, although it also occurs in Sesotho sa Leboa.

2.1 Lemmatisation of the whole word

From the following examples, it can be seen that words such as *musadzi* in Tshivenḁa, *mosadi* in Sesotho sa Leboa and *wansati* in Xitsonga have been lemmatised as whole words.

- (a) Tshivenḁa: the word *musadzi* 'woman' in Van Warmelo (1989: 235), Wentzel and Muloiwa (1982: 43), and TETD (2006: 48)
- (b) Sesotho sa Leboa: the word *mosadi* 'woman' in Kriel and Van Wyk (1983: 172)
- (c) Xitsonga: the word *wansati* 'woman' in Cuenod (1967: 224)

Every language has its own way of writing words. In other words, Tshivenḁa, Sesotho sa Leboa and Xitsonga have a disjunctive way of writing, as shown in the following sentences, e.g.

Tshivenḁa: *Musadzi u amba na niwana* 'The woman talks to the child'

Sesotho sa Leboa: *Mosadi o bolela le ngwana* 'The woman talks to the child'

Xitsonga: *Wansati u vulavula na n'wana* 'The woman talks to the child'

In the Nguni languages, a conjunctive way of writing is traditional, as shown in the following sentences, e.g.

IsiZulu: *Umfazi ukhuluma nomntwana* 'The woman talks to the child'

The preceding examples indicate that in Tshivenḁa, Sesotho sa Leboa and Xitsonga, there are five words in a sentence, including morphemes, whereas in the Nguni languages there are only three words. This is seen in Van Warmelo (1989: 235 and 289), where the verb stem *-amba* 'talks' and nouns *musadzi* 'woman' and *niwana* 'child' are lemmatised as individual words. This is also found in Wentzel and Muloiwa (1982: 7, 43, 58). In the preceding isiZulu example, however, the sentence is formed by only three words. All these words have been lemmatised as stems, as can be seen in Doke et al. (1971: 201 and 609) where the nouns *umfazi* and *nomntwana* are lemmatised as *-fazi* and *-ntwana* respectively.

The conjunctive way of writing results in lemmatisation using stems in the Nguni languages, whereas the disjunctive way of writing brings about lemmatisation of nouns using the whole word. The difference between the two ways of lemmatisation lies in the use of a hyphen before the stem, while the word can stand on its own.

Although a dictionary is not a grammar book, it includes the morphology of words because the dictionary cannot be used without some knowledge of their morphology. This is why users should be aware that at some point they would have to be able to find the stems of verbs, nouns and adjectives of the

specific language, so that they can utilise the dictionary. It is also because compilers of dictionaries use morphology when writing their dictionaries, which indicates the correct way of constructing words. This method provides the meaning of such words.

When the three dictionaries by Wentzel and Muloiwa (1982), Van Warmelo (1989), and TETD (2006) were examined, it was found that they have lemmatised the nouns using the complete word and not just the stems.

2.2 Lemmatisation of singular and plural nouns

Nouns can be lemmatised in singular or plural form. In Tshivenda, there are different class noun prefixes, which indicate the singular or plural. Such prefixes are affixed to the noun stem, as shown in the following examples:

- (a) *mushumi* 'worker' < *mu-* + *-shumi* in Van Warmelo (1989: 237),
- (b) *musidzana* 'girl' < *mu-* + *-sidzana* in Van Warmelo (1989: 237),
- (c) *vhashumi* 'workers' < *vha-* + *-shumi* is not lemmatised in Van Warmelo (1989),
- (d) *mushumi* 'worker' < *mu* + *-shumi* in TETD (2006: 49),
- (e) *vhashumi* 'workers' < *vha-* + *-shumi* in TETD (2006: 83),
- (f) *musidzana* 'girl' < *vha-* + *-sidzana* in TETD (2006: 49),
- (g) *hasidzana* 'girls' < *vha-* + *-sidzana* in TETD (2006: 83),
- (h) *mushumi* 'worker' < *mu-* + *-shumi* in Wentzel and Muloiwa (1982: 43), and
- (i) *musidzana* 'girl' *mu-* + *-shumi* in Wentzel and Muloiwa (1982: 44).

When the preceding examples are considered, it can be observed that Van Warmelo (1989) and Wentzel and Muloiwa (1982) have lemmatised the noun *mushumi* 'worker' and *musidzana* 'girl' in the singular only; whereas TETD (2006) has lemmatised both *musidzana* and *mushumi* in the singular and *vhasidzana* and *vhashumi* in the plural form. For learners who are not mother-tongue speakers of Tshivenda it will be problematic when they come across the nouns *vhashumi* 'workers' and *vhasidzana* 'girls' but cannot find them in the dictionaries.

TETD tried to solve this problem by lemmatising nouns in the singular and plural forms, as shown above. However, they did not lemmatise other Tshivenda nouns in their plural forms, e.g. *mulilo* 'fire', is lemmatised in TETD (2006: 48). However, its plural, *mililo* 'fires', is not. Also, *mulambo* 'river', is lemmatised in TETD (2006: 47). However, its plural *milambo* 'rivers', is not.

The problem here is that there is no consistency in the procedure of the compilers of these dictionaries, because some words have been lemmatised in the singular and plural, and others in the singular only.

In Tshivenda, there are 21 noun class prefixes in the singular and plural forms. Among these prefixes, some are infinite and some locative class prefixes. The noun class prefixes are as follows:

Class	Singular	Class	Plural
Class 1	<i>mu-</i>	Class 2	<i>vha-</i>
Class 3	<i>mu-</i>	Class 4	<i>mi-</i>
Class 5	<i>li-</i>	Class 6	<i>ma-</i>
Class 7	<i>tshi-</i>	Class 8	<i>zwi-</i>
Class 9	<i>n-</i>	Class 10	<i>dzi-</i>
Class 11	<i>u-</i>	Classes 10/14	<i>dzi-/vhu-</i>
Class 14	<i>vhu-</i>	Class 6	<i>ma-</i>
Class 15	<i>U</i>	Infinite noun class	
Classes 16/17/18	<i>fha-/ku-/mu-</i>	Locative noun classes	
Class 20	<i>ku-</i>	Class 8	<i>zwi-</i>
Class 21	<i>ḡi-</i>		

A noun in Tshivenda is formed by two morphemes; namely the noun class prefix plus the noun stem. The singular or plural prefix indicates the singular or plural form. In addition, every noun in Tshivenda is categorized under these classes, e.g.

Singular	Plural
<i>musadzi</i> 'woman'	> <i>vhasadzi</i> 'women'
<i>muthu</i> 'person'	> <i>vhathu</i> 'persons'
<i>mulilo</i> 'fire'	> <i>mililo</i> 'fires'

The examples above show that the noun *musadzi* falls under the prefix *mu-* in the singular and under the prefix *vha-* in the plural *vhasadzi*. The noun *mulilo* falls under the singular class prefix *mu-*, whereas it comes under the plural prefix *mi-* in *mililo*. Among these nouns there are instances where the prefix is not visible in the singular form whereas in the plural it is, e.g.

Singular	Plural
<i>mme</i> 'mother' class 1a	<i>vhomme</i> 'mothers' class 2b
<i>khotsi</i> 'father' class 1a	<i>vhokhotsi</i> 'fathers' class 2b

In the preceding examples, it may appear that there is no class prefix attached to the nouns. What is seen here are the stems *mme* and *khotsi*. To these stems, which are recognized nouns, the plural prefix *vho-* can be affixed to form the plural nouns *vhomme* and *vhokhotsi*. This prefix *vho-* indicates that the noun to which it is affixed is in the plural form. These words should be interpreted as two different words. Lemmatising nouns using a singular or a plural prefix started some time ago; for example, Van Warmelo (1989) and Wentzel and Muloiwa (1982). In Wentzel and Muloiwa (1982: 43) lemmatisation has been

done as follows:

musadzi (*vha-*)

This shows that the noun *musadzi* is in the singular form, and its plural form is indicated as *vha-* in brackets. Van Warmelo (1989: 235) has not indicated the plural of the noun *musadzi*. He has given it as follows in the singular form only:

musadzi 1 (Class 1)

This merely shows that the noun *musadzi* falls under the first class prefix of nouns, and is in the singular form. Not all nouns in the plural form are lemmatised in this way; others are represented as follows:

Miḍali is the plural form of *muḍali*

mililo is the plural form of *mulilo*

These nouns appear to have been lemmatised in Van Warmelo (1989: 196 and 197) in alphabetical order. However, *muḍali* in the singular and *miḍali* in the plural form have been omitted in TETD (2006). The difference between *mulilo* and *mililo* is brought about by the class prefix *mu-* of the singular, meaning that the word refers to a single entity, while the plural class prefix *mi-* indicates that the word refers to more than one thing. This way of lemmatisation on the basis of the singular and plural forms has advantages and disadvantages.

2.2.1 Advantages of the lemmatisation of singular and plural nouns

It is user-friendly when all nouns will be lemmatised in the singular and the plural. This has been emphasised by Prinsloo and De Schryver (1999: 267) as follows:

The major advantage of this lemmatization procedure is user-friendliness rendering a practical theory.

With this way of lemmatisation, the dictionary is easy to use, because no intensive knowledge of the language is expected from users. What is required is just knowledge of the alphabet of Tshivenda. When we look at the Tshivenda noun *mushumi* 'worker/labourer', it is lemmatised in the singular by Van Warmelo (1989: 237). He has also lemmatised the plural form of *mushumi* 'worker/labourer', which is *vhashumi* 'workers/labourers'. This is a user-friendly format even for a person who is not *Muvenda*.

Wentzel and Muloiwa (1982: 43) also lemmatised the noun *mushonga* 'medicine', and the plural for *mushonga* 'medicine' is shown by the prefix *mi-*. By doing this, he is indicating that when one wants to get the plural form of *mushonga* 'medicine' one should just add the prefix *mi-* to the noun stem *-shonga* to make it *mishonga* 'medicines'. This is a user-friendly approach.

2.2.2 Disadvantages of the lemmatisation of singular and plural nouns

There are also problems associated with the lemmatisation of both singular and plural nouns. Firstly, because plurality appears in all Tshivenḁa nouns, compilers of dictionaries will have to pose the question: Which nouns should be selected for lemmatisation? As a result compilers often fail to lemmatise important, frequently occurring nouns, thus rendering the dictionary ineffective to its users.

The following nouns have been lemmatised in the singular form in Van Warmelo (1989: 237 and 235):

musidzana 'girl'
musadzi 'woman'

However, the following plurals of these nouns have not been lemmatised:

vhasidzana 'girls'
vhasadzi 'women'

The dictionary seems to have been written only for users who are conversant in Tshivenḁa; that is users who would know that the plural of the noun *musidzana* is *vhasidzana*. By lemmatising the nouns in the singular only, the plurals of important nouns in Tshivenḁa are left unlemmatised.

In Wentzel and Muloiwa (1982: 42 and 44) the following nouns have been lemmatised in both the singular and plural, and the plural is indicated by *vha-* as shown below:

munna (*vha-*)
musidzana (*vha-*)

This means that the noun *musidzana* is the singular form, and to form the plural the prefix *vha-* must be added. This way of lemmatisation is very effective for those who know Tshivenḁa. However, users who do not know Tshivenḁa cannot consult the dictionary to check plurality. When they come across *vhasidzana*, they would not know that it is the plural of *musidzana*. When they come across the plural *vhasidzana*, instead of the singular *musidzana*, the users would be unaware that it is the plural of *musidzana*. It therefore renders the dictionary ineffective to users who are not conversant in Tshivenḁa. In this way, dictionaries omit important words in the plural.

2.3 Lemmatisation of deverbative nouns

Poulos (1990: 14) defines *deverbative nouns* as follows:

A number of nouns in some of the classes are derived from other parts of speech, for example, from verbal radicals (or verb roots).

In Tshivenda, the derivation of nouns from verb stems or adjective stems occur by affixing singular and plural prefixes to these stems and suffixes or endings *-i, -e, -o, -u*.

Nouns	derived from verbs
<i>ndima</i>	<i>-lima</i> 'to plough'
<i>murengi</i>	<i>-renga</i> 'to purchase'
<i>mulimi</i>	<i>-lima</i> 'to plough'
<i>muimbi</i>	<i>-imba</i> 'to sing'
<i>tshibiki</i>	<i>-bika</i> 'to cook'
<i>vhuhulu</i>	<i>-hula</i> 'to be big'
<i>vhusekene</i>	<i>-sekene</i> 'to be thin'
<i>vhulapfu</i>	<i>-lapfa</i> 'to be long'

(Nthambeleni 2008: 49)

Other deverbatives formed from verbs with suffixes or endings added are the following:

<i>murungo</i>	'seam'
<i>muvhuso</i>	'government'

Because they are problematic, the preceding nouns should have been lemmatised in the dictionary. Compilers should include deverbatives like these in their dictionaries.

In African languages, nominal prefixes are therefore important to such an extent that prefixes of class nouns and suffixes or endings such as *-a, -e, -i, -o,* and *-u* should be considered when deverbatives are discussed.

Nouns formed in this manner are not lemmatised in great numbers in dictionaries, as is the case in Tshivenda dictionaries. This results in ineffective dictionaries being produced, which disadvantages dictionary users. Lemmatisation of these words in the dictionary depends on the way speakers use them. When they are used in language, compilers of dictionaries should select them for lemmatisation.

Focusing again on the above-mentioned items, compilers of Tshivenda dictionaries have fulfilled an important function in lemmatising deverbatives. However, it is problematic when some deverbative nouns have not been lemmatised as in Van Warmelo (1989: 52). Here the verb stem *-fhaṭa* 'to build' has been given a deverbative noun by affixing the class prefix *mu-* to a verb which is not given. This shortcoming causes problems for non-native users of the dictionary. Such users will find the noun *muḥḥaṭi* 'builder' but not know its meaning. This is also the case with the noun *murengi* 'buyer'. Although they are used frequently in Tshivenda, these nouns have not been lemmatised.

2.4 Lemmatisation of diminutive nouns

Poulos (1990: 81) defines *diminutives* as follows:

Those expressions in a language that basically express the idea of 'smallness', 'shortness' and where appropriate the 'young of' some or other noun.

As users would look up diminutives in the dictionary, it is important to lemmatise them. In Tshivenḡa, diminutives can be formed by affixing the different class prefixes and suffixes to the nouns, such as the suffixes *-ana*, *-nyana*; and class prefixes 7 *tshi-*, 11 *lu-*, 20 *ku-* in the singular and class 8 *zwi-* and 14 *vhu-* in the plural, as in the following examples (not from the dictionaries mentioned above).

Class 7 *tshi-* + *-budzi* > *tshibudzi* (small fat goat)

Class 20 *ku-* + *-budzi* > *kubudzi* (small young goat)

Each of these nouns *tshibudzi* and *kubudzi* has a different meaning. They are therefore lemmatised in dictionaries.

When the suffix *-ana* is added to nouns, it forms diminutive nouns

mbudzi + *-ana* > *mbudzana*

The suffix *-nyana* also forms diminutives and, when added to nouns, it results in the following:

- (a) When it is added to nouns it may refer to a liquid which is little in quantity.
 - (i) *maḡi* + *-nyana* > *maḡinyana* (little water)
 - (ii) *mafhi* + *-nyana* > *mafhiniyana* (little milk)
 - (iii) *malofha* + *-nyana* > *malofhanyana* (little blood)
- (b) When it is used with a class prefix and a verb stem, it may indicate that there is not much of something.
 - (i) *mu* + *-shum* + *-o* + *-nyana* > *mushumonyana* (little work)
- (c) When it is added to nouns, it may indicate smallness in size.
 - (i) *musidzana* + *-nyana* > *musidzanyana* (small girl)

The noun prefix plus suffix function together when added to the noun stem to form a diminutive noun, as in the following examples:

tshi- + *-budz-* + *-ana* > *tshibudzana* (small goat)

zwi- + *-budz-* + *-ana* > *zwibudzana* (small goats)

vhu- + *-budz-* + *-ana* > *vhubudzana* (thin goats)

From the examples shown above, it becomes evident that, just like other nouns, diminutives are important in a language. In addition, in a language like Tshivenḡa, they are used frequently.

When the different dictionaries in Tshivenda are examined, one will find that they have lemmatised diminutive nouns. The noun *mbudzi* (goat) has been lemmatised, whereas its diminutive form *tshibudzana* (small goat) has not been listed in Van Warmelo (1989), although it is used frequently. Omitting it from the dictionaries is problematic for the users who do not know Tshivenda because they would not be able to find the word *tshibudzana* and its meaning.

When nouns are looked up in Tshivenda dictionaries such as Wentzel and Muloiwa (1982), Van Warmelo (1989) and TETD (2006), it becomes clear that several diminutive nouns have not been lemmatised. Compilers should ensure that all such nouns are lemmatised. Only when this is done, will good dictionaries be compiled that will satisfy users' needs.

3. Conclusion

When the current Tshivenda dictionaries are examined, it becomes clear that all nouns should be lemmatised. The main aim of lemmatisation of nouns is to enable compilers of dictionaries to produce useful dictionaries, and not repeat the same mistakes made in previous dictionaries. For the production of useful dictionaries that will benefit users, nouns should be lemmatised in a well-considered manner. This will also support De Schryver and Prinsloo's (2000) idea of 'Simultaneous Feedback', since a dictionary should be written in such a way that users can find information easily. It has been established that Tshivenda nouns are lemmatised in both the singular and plural forms; as diminutives and deverbatives. Including nouns in singular and plural form, as well as diminutive and deverbative nouns, should be formulated as a lexicographic principle. Failure to meet this requirement would render the dictionary not user-friendly at all.

References

- Chambers-Macmillan.** 1996. *South African Student's Dictionary*. Manzini: Macmillan Boleswa.
- Cuenod, R.** 1967. *Tsonga-English Dictionary*. Braamfontein: Sasavona.
- De Schryver, G.-M. and D.J. Prinsloo.** 2000. The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. *Lexikos* 10: 1-31.
- Doke, C.M. et al.** 1971. *English-Zulu Dictionary*. Johannesburg: Witwatersrand University.
- Gove, P.B. (Ed.)**. 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: Merriam-Webster.
- Hartmann, R.R.K. and G. James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- Hartshorne, K.B.** 1984. *Dictionary of Basic English: Venda*. Assisted by J.H.A. Swart and E. Posselt. Johannesburg: Educum.
- Joubert, P.J. and D.N. Rapea.** 1972. *Verklarende Woordelys/Mutevhe wa Maipfi o Țalutshedzwaho: Afrikaans-English-Venda/Tshivuhuru-Tshiisimane-Tshivoenda. Vol.1 A-C*. SABC.
- Kriel, T.J. and E.S. van Wyk.** 1983. *Pukuntšu Dictionary*. Pretoria: J.L. van Schaik.

Macmillan English Dictionary for Advanced Learners. (Edited by M. Rundell and G. Fox.) 2002. London: Macmillan.

Marole, L.T. 1955. *Afrikaans–Venda: Vocabulary and Phrase Book*. Sibasa: Marole Book Depot.

Marole, L.T. and F.J. de Gama. 1954. *English–Venda Vocabulary*. Third edition. Sibasa: Marole Book Depot.

Nthambeleni, M. 2008. *Thaidzo musi hu tshi Lemathaizwa Madzina na Maiti kha Ṭhalusamaipfi dza Tshivenda*. Unpublished M.A. Dissertation. Thohoyandou: University of Venda.

Plisson, J., D. Mladenić, N. Lavrač and T. Erjavec. 2005. A Lemmatization Web Service Based on Machine Learning Techniques. Vetulani, Z. (Ed.). 2005. *Proceedings of the 2nd Language & Technology Conference, April 21–23, 2005, Poznań, Poland*: 369-372. Poznań: Wydawnictwo Poznańskie.

Poulos, G. 1990. *A Linguistic Analysis of Venda*. Pretoria: Via Afrika.

Prinsloo, D.J. and G.-M. de Schryver. 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *Southern African Journal of African Languages* 19(4): 258-275.

Sinclair, J.M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tshivenda National Lexicography Unit. 2006. *Tshivenda/English Ṭhalusamaipfi/Dictionary*. Cape Town: Phumelela Publishers.

Van Warmelo, N.J. 1989. *Venda Dictionary: Tshivenda–English*. Pretoria: J.L. van Schaik.

Wentzel, P.J. and T.W. Muloiwa. 1982. *Ṭhalusamaipfi ya nyambotharu yo khwiniswaho: Luwenda–Luwuhuru–Luisimane/Verbeterde drietalige woordeboek: Venda–Afrikaans–Engels/Improved Trilingual Dictionary: Venda–Afrikaans–English*. Pretoria: University of South Africa.

Elicitation and Arrangement of Conceptual Meanings in the Lexicography of Less Documented Languages

Celestino Oriikiriza, *Department of Linguistics, English Language Studies
and Communication Skills, Makerere University, Kampala, Uganda*
(oriikiriza@chuss.mak.ac.ug/oriikiriza@yahoo.co.uk)

Abstract: The paper demonstrates how some of the dictionaries written in less documented languages, hardly meet the expectations of target users, due to some of the methods used in collecting and arranging meanings of words. The paper, therefore, explains the semantico-syntactic method of eliciting multiple conceptual meanings of words and the alphasyntactico-semantic mode of their arrangement in dictionary making. It concludes by showing how the two methods can lead to compilation of good dictionaries in less documented languages and how the dictionaries would be of benefit to the target users.

Keywords: ARRANGEMENT OF MEANINGS, CONCEPTUAL MEANINGS, DICTIONARY, LESS DOCUMENTED LANGUAGES, LEXICOGRAPHY

Opsomming: Onthulling en ordening van konseptuele betekenis in die leksikografie van minder gedokumenteerde tale. In hierdie artikel word aangetoon dat woordeboeke vir tale wat nie goed gedokumenteer is nie, dikwels nie voldoen aan die verwagtinge van die teikengebruikers nie as gevolg van sommige metodes wat gebruik word in die versameling en ordening van betekenisonderskeidinge. Gevolglik word 'n uiteensetting gegee van die semanties-sintaktiese metode oor hoe om veelvuldige betekenisonderskeidinge van woorde te onthul en van die alfa-sintakties-semantiese metode oor hoe die ordening in woordeboeke kan geskied. Ter afsluiting word aangetoon hoe hierdie twee metodes tot die samestelling van goeie woordeboeke kan lei vir tale wat nie goed gedokumenteer is nie en hoe hierdie woordeboeke die teikengebruikers kan bevoordeel.

Sleutelwoorde: BETEKENISORDENING, KONSEPTUELE BETEKENIS, LEKSIKOGRAFIE, MINDER GEDOKUMENTEERDE TALE, WOORDEBOEK

1. Introduction

Dictionaries: definition and importance

A dictionary is a book in which words are listed alphabetically and their meanings, either in the same language or in another, and other information

about them are given (Procter et al. 1995). Dictionaries shape the language, for instance, by presenting the standard form of a language to the dictionary user. They are used for reference in terms of meaning, spelling, pronunciation, grammar, synonyms, choice of words and for general information (e.g. geographical names and units of measure). Dictionaries are also language inventories. This paper traces the reasons why some of the dictionaries do not quite satisfy the user's needs on the particular aspect of meanings. It presents the semantico-syntactic method of elicitation of meanings of words and the alphasyntactico-semantic mode of arrangement of meanings as a solution to the problem.

2. The Formal Stages of Compiling a Dictionary

Cases abound where some language specialists or even enthusiasts write dictionaries without going through the formal stages of dictionary writing. In such cases, they collect words, and start defining their meanings. However, collection of words and definition of meanings are only part of the several formal stages of writing a dictionary. There is no doubt that a dictionary will achieve its objectives, if it is compiled following the formal stages of writing a dictionary.

Landau (1989) and Singh (1991) present the formal stages of compiling a dictionary. Broadly, the stages are: planning the dictionary, writing the dictionary and producing the dictionary. Revising the dictionary and abridging the dictionary are secondary stages of writing a dictionary. There are sub-phases under each of the steps. The stage of writing the dictionary is pertinent to a lexicographer who is writing a dictionary against the background of the need to record most basic meanings of words. Depending on the type of a dictionary (e.g. a general-purpose dictionary), the lexicographer needs to use a dictionary style manual that matches the intended goal, particularly on the method of eliciting multiple meanings of words and the arrangement of the meanings in a consistent order.

3. Limitations of Meaning Elicitation and Arrangement in Dictionaries

Taking the case of Uganda, the following backgrounds have influenced the writing of different dictionaries:

- (i) Feelings of identity by speakers of minority languages, e.g., Lugungu language that is currently being recorded in a dictionary form.
- (ii) The need to preserve the language, as in the case of Runyankore-Rukiga, which has had several dictionary projects.
- (iii) The introduction of the thematic school curriculum. Uganda launched a school programme whereby mother tongues are used as mediums of in-

struction for pupils in the lower primary school (P1–P3). The pupils learn about different topics, from which aspects of science, maths, social studies and grammar are highlighted. It is partly for ease of comprehension, because children learn better in the language they understand. It is also partly for building a foundation for the teaching of Ugandan languages in school, as plans are underway by the government to offer the languages as subjects of study in upper levels of school. In the upper levels, the indigenous languages are supposed to be studied alongside foreign languages, particularly English. Given this policy, some publishing houses have responded by writing and publishing dictionaries of various types in the approved languages. The dictionaries are written upon request by the government and others according to the good judgement of the publishing house.

- (iv) Missionary activities: Missionary activities in Uganda saw the writing of different dictionaries in some local languages. The dictionaries were mainly bilingual for purposes of language learning by the missionaries and colonial agents, particularly educationists. For example, in 1917 and 1959, two dictionaries were published, namely, a Luganda–French dictionary and a Runyankore–Rukiga–English and English–Runyankore–Rukiga dictionary written by Le Veux and Taylor respectively.

The objectives and uses of the dictionaries cited above are good because they aim at the documentation of languages, finding alternative ways of mediums of instruction and solving the communication gap.

However, the most intriguing problem comes when such dictionaries cannot fulfil the expectations of the users in schools, or in ordinary situations, and if they cannot be judged as proper reflections of what is in the language. This is what Kiango (2000: 4) alludes to by noting that *early dictionaries compiled by the missionaries and the colonial administrators, were not compiled with a native speaker in mind as the prime user ... these dictionaries were not aimed at being complete records of the languages concerned and up to this time no such records exist.*

With respect to meanings and their arrangement in a dictionary, a comparative analysis of a Ugandan language dictionary with an English dictionary (published by a reputable house), results into noticeable differences. The dictionaries in English tend to have multiple meanings of words than dictionaries in most Ugandan languages. They also tend to have meanings of words arranged in a consistent order than their counterparts referred to above. The difference is caused by the methods used for generating meanings of words and arranging the meanings in a consistent order.

Corpus-driven methods can generate multiple meanings. They can also lead to consistent arrangement of the meanings in dictionary form than field-work-based methods. However, corpus-driven methods easily work in well-documented languages where written corpus is readily available than in less documented languages where it is not. For example, the level of documentation of Ugandan languages is very low compared to that of English.

As a result, dictionaries in well-documented languages are much more likely to have a higher coverage of meanings of words than dictionaries in less documented languages. Oriikiriza (2011) cites an example of *make* (v.) in several English dictionaries; *okukora* (v.), a Runyankore-Rukiga word for 'work', 'do', 'make' in some Runyankore-Rukiga dictionaries; and *la* (v.), a Kiswahili word for 'eat' in some Kiswahili dictionaries. The work indicates that Summers et al. (2003), an advanced learner's dictionary, has 27 main meanings of the word *make*, and Soanes (2001), which is a pocket-size volume, has 10 main meanings of the word. In the case of the Runyankore-Rukiga example *okukora* (v.), the work indicates that the latest and standard dictionary among the ones that are cited for Runyankore-Rukiga is Oriikiriza (2007). It is described as a general-purpose dictionary with 4 main meanings of *okukora* (v.). For Kiswahili, TUKI (1981), the work describes it as a general-purpose Kiswahili dictionary, and that it records 8 main meanings for *la*. Therefore, basing on these examples, a general-purpose dictionary in a Bantu language (e.g. Runyankore-Rukiga and Kiswahili) has a coverage of meanings that is equivalent or almost equivalent to that of a pocket-size English dictionary. This difference is attributed to the methods of obtaining the meanings.

Use of informants (fieldwork) and the lexicographer's knowledge of the language are the main methods of generating meanings of words and arranging them in dictionary form in less documented languages, e.g. Bantu, as compared to the corpus-driven methods in well-documented languages. In the former case, the methods rely on memory and so are prone to situations where the informant or lexicographer cannot recall most of the meanings of a word. Secondly, the informant or the lexicographer sequences meaning of words according to the order in which he/she knows them. This sequence is subjective since the dictionary target users do not necessarily store meanings of words in the same order according to their innate knowledge of the meanings. In order to solve the problem, the corpus-driven methods of generating meanings of words and arranging meanings in a consistent order would substitute the fieldwork methods. However, they are affected by lack of sufficient written materials to generate the corpus. In this regard, there is need to develop other non-corpus-driven methods that can be used. One of the methods being proposed in this paper is the semantico-syntactic method of elicitation of conceptual meanings of words and the alphasyntactico-semantic mode of arrangement of the meanings.

4. The Semantico-syntactic Method of Elicitation of Conceptual Meanings of Words and the Alphasyntactico-semantic Mode of Arrangement of the Meanings

4.1 Classification of Meanings

Lexical meaning can be defined as the sense, reference, usage and utterance of

a word. Sense in this context refers to the basic meaning of a word, while reference refers to the extended meanings. The usage meaning is associated with the social usage of the word. These kinds of meanings can be explained using an example of *sit* (v.), 'to rest or relax on a chair or seat'. As a basic meaning, it is therefore the sense. In a sentence such as *The committee sits tomorrow*, *sit* means 'to hold a meeting'. This is an extended meaning derived from the sense. It should be noted that words may have meanings resulting from usage according to region, medium, attitude, status, field (of practice, e.g. legal), temporal perspective (e.g. obsolete, archaic), situation (e.g. formal, informal, elevated style, slang), etc. These are usage meanings. They may also have meanings resulting from the speech act with which the word is used. The meanings in this case are referred to as utterance meanings.

Taking the example of *sit*, Summers et al. (2003) records as British English (i.e. regional/geographical usage) its meaning of 'taking an exam'. It also records it as having a specific imperative meaning, 'when used to tell a dog to sit in an upright position'. A command in this case is an utterance meaning.

For sense and reference meaning in particular, it suffices to say that sense results from the basic pattern with which a word is used and reference meaning results from a different entity with which a word is used. Therefore, the basic pattern of *sit* is [something sits on something], e.g., *The child is sitting on a mat*. In this pattern, *sit* denotes the basic meaning, i.e. 'resting your body in a chair or seat'. It gets a reference meaning when the entities in the same pattern are varied. For example, in *The building is sitting on an acre piece of land*, *sit* means 'to be positioned or found in a particular place'. This sentence is of the same pattern as the one above, but the entities being interrelated by the verb, namely 'building' and 'acre piece of land' are different from those in *The child is sitting on a mat*. The variation in entities results into a different meaning of the verb.

All these kinds of meanings outlined above need to be compiled in a dictionary, particularly a general-purpose dictionary. The sense and reference meanings are collectively known as conceptual meanings, while the usage and speech act meanings are known as interpersonal meanings. This paper only discusses the elicitation and arrangement of conceptual meanings (sense and reference), using the semantico-syntactic method and the alphasyntactico-semantic mode respectively (see 4.3 below). The semantico-syntactic method of elicitation of meanings of words is based on the situation-role theory of meaning (Kiingi 2008; Kiingi 2009). The method can be used to elicit conceptual meanings of words, i.e. senses and reference meanings. The alphasyntactico-semantic mode of arrangement of the meanings is partly based on the situation-role theory, and partly on the alphabetical mode of arrangement. Both methods will be exemplified later on; however, the situation-role theory from which they are derived is explained below.

4.2 The Situation-Role Theory of Meaning

The situation-role theory of meaning is a modified version of the semantic-role theory of meaning. The theory postulates the following: situations, semantic categories and semantic-roles.

4.2.1 Situation

A situation is a state-of-affairs expressed in a sentence.

4.2.2 Semantic categories

Semantic categories are entities that perform the roles in the situations. Table I provides an outline of these categories and their corresponding examples of entities.

Table I: Semantic Categories

GENERAL SEMANTIC CATEGORY (IN NAME AND ABBREVIATION)	EXAMPLES OF ENTITIES IN THE CATEGORY
(1) human (h)	Peter, John, he, she, they
(2) animate (b)	dog, giraffe, institution or organisation
(3) concrete (r)	chair, pot, car, key, door, plant, animal, building
(4) event (e)	write, dance, beg, open, study, pay
(5) abstract (a)	light, sound, heat, energy, a non-physical report as opposed to a physical (i.e. concrete) report, and a school (as an institution)
(6) state, i.e. quality (q)	happy, sick, sad
(7) quantity, i.e. group (g)	sets of things, e.g. team, pair of shoes, group
(8) number (n)	numerical objects, e.g. once a year, twice in a week, forty days, two billion
(9) space (l)	spaces, e.g. up, down, inside, town, Kampala, London
(10) time (t)	temporal objects, e.g., today, tomorrow, yesterday, a fortnight ago

4.2.3 Semantic roles

Semantic roles refer to the roles performed in situations as indicated in the Table II below.

Table II: Semantic Roles

SEMANTIC-ROLE	ABBREVIATION	EXAMPLE
(1) Volitional, Affected and Dynamic bearer	V	<i>The <u>students</u> went to class.</i>
(2) Non-volitional, Affected and Dynamic bearer	B	<i>The hunter killed <u>an animal</u>.</i>
(3) Effected: outcome or result in a situation	X	<i>He wrote <u>a letter</u>.</i> <i>He made <u>a pot</u> from clay.</i>
(4) Volitional Affected Non-dynamic bearer	W	<i><u>The driver</u> stood near the car.</i>
(5) Non-Volitional, Non-affected and Non-dynamic bearer	Q	<i><u>The clothes</u> are dry.</i> <i><u>The door</u> is open.</i>
(6) Follower: a situation that follows another	F	<i>The substance changed from a liquid to <u>a gas</u>. (Larson and Segal 1995: 482)</i>
(7) Direction: the category faced in a situation	D	<i><u>The school</u> is opposite the mosque.</i>
(8) Reference: the category referred to in a situation	R	(a) <i>They are <u>at home</u>.</i> (b) <i>The exam starts <u>at nine</u>.</i>
(9) Source: the category at the beginning of a situation	S	(a) <i>The substance changed from <u>a liquid</u> to a gas. (Larson and Segal 1995: 482)</i> (b) <i>The inheritance passed from <u>Jill</u> to Kate. (Larson and Segal 1995: 482)</i>
(10) Mediate: the category between two categories in a situation	M	<i>They travelled to London <u>by train</u>.</i>
(11) Goal: the category at the end of a situation	G	(a) <i>The substance changed from a liquid <u>to a gas</u>. (Larson and Segal 1995: 482)</i> (b) <i>The inheritance passed from Jill <u>to Kate</u>. (Larson and Segal 1995: 482)</i>
(12) Possessum: the category possessed in a situation	H	<i>We have <u>little food</u>.</i> <i>Harriet owns <u>a cat</u>. (Brown and Miller 1980: 309)</i>
(13) Comitative: the category accompanying another in a situation	J	<i>The bride walked with <u>the groom</u>.</i>

(14) Stimulus: The category perceived in a situation	T	<i>We saw <u>a dog</u>.</i> <i>We heard <u>birds</u> singing in the tree.</i>
(15) Volitional entity generator	I	<i><u>He</u> wrote a letter with a pen.</i>
(16) Non-Volitional entity generator	N	<i><u>The machine</u> makes bricks.</i>
(17) Volitional affector	Y	<i><u>They</u> elected him chairman.</i>
(18) Non-volitional affector	Z	<i><u>The stone</u> shattered the window.</i>
(19) Volitional recipient	E	(a) <i>Ali gave <u>Fatuma</u> a book.</i> (b) <i>It gave <u>us</u> a big problem.</i>
(20) Non-volitional recipient	O	<i>Ali gave <u>the door</u> a kick. (I.e. to give sth. sth.)</i>
(21) Volitional event causer	A	<i><u>Peter</u> opened the door.</i>
(22) Non-volitional event causer	C	<i><u>The key</u> opened the door.</i>

The following proofs are made from the situation-role theory:

- (i) General structures which specify the place of a semantic-role in a sentence. The structures are called well-formed formulae (wff) as shown in the first column of Table III below.

Table III: Well-formed Formulae for Semantic-Roles

Well-formed formulae (with θ [theta] standing for semantic-role and ϵ [epsilon] for semantic category)	Entities playing the semantic roles (underlined)	Sentence expressed in terms of semantic-roles and semantic categories indicated in Table I & II
$[\theta\epsilon]$	<u>The door</u> opened.	Br
$[\theta\epsilon_1 \theta\epsilon_2]$	<u>He</u> opened <u>the door</u> .	Ah Br
$[\theta\epsilon_1 \theta\epsilon_2 \theta\epsilon_3]$	<u>He</u> opened <u>the door with a key</u> .	Ah Br ₁ Rr ₂

- (ii) Semantic-Role Patterns

Just as there are (basic) sentence patterns, there are also semantic-role patterns as shown below.

Table IV: Semantic-Role Patterns

Semantic-role Pattern	Explanation and example
Ω	The symbol is read as omega. It stands as a semantic-role pattern for situations involving the following roles: V (<i>He laughed</i>), B (<i>The door opens</i> , <i>The ball reddens</i>), X (<i>The ice is melting</i>)
Ψ	The symbol is read as psi. It stands as a semantic-role pattern for situations involving the following roles: V (<i>He turned a traitor</i>), B (<i>The ball becomes red</i>), X (<i>The children are playing well</i>); W (<i>Ali is a doctor</i>), Q (<i>The ball is red</i>)
$K' \Psi$	The symbol K' is read as kei prime. It combines with psi to form a combinational pattern in which the following roles occur: a causer of a causer (K') and any of the Ψ roles. Example: <i>They painted the door red</i> . This situation presupposes 'The door is red', 'The door becomes red' and 'They made the door red'. It can be expressed in terms of the semantic-roles and semantic categories as Ah [Br Gq].
$\Psi \Phi$	The symbol Φ is read as phi. It stands with psi to form a combinational pattern involving any of the Ψ and Φ roles. The roles in the latter case are F, D, R; S, M, G; H, J, T. Example: <i>The children went to school</i> . Bh R ₁
$K \Psi$	This is a combinational pattern involving a causer (K) and any of the Ψ roles. The K roles are I, N; Y, Z; O, E; A, C. Example: <i>John killed a dog</i> . Ah Bb
$K \Psi \Phi$	This is a combinational pattern involving three semantic-roles, i.e., any of the K, Ψ and Φ . Example: <i>Jane knitted a sweater for John</i> . Ih ₁ Xr Oh ₂
$K' K \Psi$	This is a combinational pattern involving three roles, i.e., any of the K' , K and Ψ roles respectively. Example: <i>He opened the door with a key</i> . Ah Br ₁ Rr ₂

(iii) Syntactic-role patterns

The term syntactic-role patterns is akin to sentence patterns. Except that, the syntactic-role patterns do not correspond exactly to the sentence patterns. The degree of correspondence can be seen in the examples in Table V.

Table V: Degree of correspondence between syntactic-role patterns and sentence patterns

Syntactic-role Patterns	Sentence Patterns
S (Subject): <i>The door opens</i> .	SV (Subject + Verb): <i>The door opens</i> .
SN (Subject + Neutral element): <i>The ball is red</i> .	SVC (Subject + Verb + Complement): <i>Ali is a doctor</i> .

SCN (Subject + Constrained object + Neutral complement): <i>They painted <u>the door red</u>.</i>	SVA (Subject + Verb + Adverbial): <i>The children went <u>to school</u>.</i>
SF (Subject + Free object): <i>The children went <u>to school</u> / He saw <u>a dog</u>.</i>	SVO (Subject + Verb + Object): <i>He killed <u>a dog</u>.</i>
SC (Subject + Constrained object): <i>He killed <u>a dog</u>.</i>	SVOC (Subject + Verb + Object + Complement): <i>They elected <u>him president</u>.</i>
SCF (Subject + Constrained object + Free object): <i>Jane knitted <u>a sweater for John</u>.</i>	SVOA (Subject + Verb + Object + Adverbial): <i>Jane knitted <u>a sweater for John</u>.</i>
SCC (Subject + Constrained object + Constrained object): <i>He opened <u>the door with a key</u>.</i>	SVOO (Subject + Verb + Object + Object): <i>He opened <u>the door with a key</u>.</i>

There is a mismatch in the placement of the situations. What is a complement in the sentence patterns is a neutral element in the syntactic-role patterns, if it expresses a state, or quality. This is one of the differences between syntactic-role patterns and sentence patterns.

(iv) Semantico-syntactic Isomorphism

The explanation above systematises syntactic-role patterns and semantic-role patterns. The juxtaposition below shows that there is a one-to-one correspondence between semantic-role patterns and syntactic-role patterns, signified by the symbol \cong . In this way, the patterns can be called semantico-syntactic patterns, or syntactico-semantic patterns:

- (1) $[\Omega] \cong S$
- (2) $[\Psi] \cong SN$
- (3) $[K' \Psi] \cong SCN$
- (4) $[\Psi \Phi] \cong SF$
- (5) $[K \Psi] \cong SC$
- (6) $[K \Psi \Phi] \cong SCF$
- (7) $[K' K \Psi] \cong SCC$

(v) Phrasal Categorial Patterns

Semantico-syntactic patterns can be expressed in terms of phrasal categorial patterns as demonstrated in Table VI.

Table VI: Phrasal Categorial Patterns

Semantic-role pattern	Examples of a semantic-role playing entity (underlined)	Phrasal-categorial patterns of the entities
Ω	<i>The <u>door</u> opens.</i>	N" (Noun Phrase)
Ψ	<i><u>Ali</u> is a doctor. The <u>door</u> is open.</i>	N" N" (Noun Phrase, Noun Phrase) N" A" (Noun Phrase, Adjective Phrase)

K' Ψ	<i>They elected <u>him</u> president.</i> <i>They painted <u>the door</u> red.</i>	N" N" N" (Noun Phrase, Noun Phrase, Noun Phrase) N" N" A" (Noun Phrase, Noun Phrase, Adjective Phrase)
Ψ Φ	<i>He came <u>yesterday</u>.</i> <i>The cat is <u>under the table</u>.</i>	N" Adv" (Noun Phrase, Adverb Phrase) N" P" (Noun Phrase, Prepositional Phrase)
K Ψ	<i>He opened <u>the door</u>.</i> <i>He was hit <u>by a bullet</u>.</i> <i>He came <u>to show me the report</u>.</i>	N" N" (Noun Phrase, Noun Phrase) N" P" (Noun Phrase, Prepositional Phrase) N" S' (Noun Phrase, Clause)
K Ψ Φ	<i>He wrote <u>the report well</u>.</i> <i>He gave <u>the book to me</u>.</i>	N" N" Adv" (Noun Phrase, Noun Phrase, Adverb Phrase) N" N" P" (Noun phrase, Noun Phrase, Prepositional Phrase)
K' K Ψ	<i>He gave <u>me a book</u>.</i> <i>The venture caused <u>me trouble</u>.</i> <i>He opened <u>the door with a key</u>.</i> <i>They elected <u>him as their representative</u>.</i>	N" N" N" (Noun Phrase, Noun Phrase, Noun Phrase) N" N" P" (Noun Phrase, Noun Phrase, Prepositional Phrase) N" N" S' (Noun Phrase, Noun Phrase, Clause)

4.3 Application of the Situation-Role Theory in Meaning Elicitation and Arrangement

The corpus-driven method is the conventional approach for elicitation of meanings of words. It is also the conventional approach for the arrangement of meanings of words, particularly in the arrangement of meanings of words by order of their frequency. However, it is not quite viable in the dictionary compilation of less documented languages, since it relies on sufficient written materials that are not easy to find in a less documented language. Oriikiriza (2011) argues that multiple meanings can be easily elicited and consistently arranged using the semantico-syntactic method and the alphasyntactico-semantic mode respectively, without the written materials of a language. The methods are mainly based on the Situation-Role Theory of meaning (Kiingi 2008; Kiingi 2009), and largely depend on the grammatical structure of language. They are stated as follows:

STEP 1:

For a given word, find out a sentence (i.e. situation) in which its basic (i.e. core) meaning is applied. The sentence is meant to enable us to predict other meanings of the target word in STEP 2 and 3.

STEP 2a:

Find the semantic-role patterns with which the word is used, by constructing sentences based on the ascending complexity of syntactico-semantic patterns.

STEP 2b:

Using semantic category switching and semantic category variation, construct more sentences.

STEP 3a:

Formalise the sentences (i.e. express the sentences in terms of semantic-roles and semantic categories).

STEP 3b:

Investigate the meaning of the word by looking at each of the example sentences in terms of its formalisation.

STEP 4:

Order the meanings according to:

- (i) ascending complexity of semantico-syntactic patterns (i.e. valency);
- (ii) ascending complexity of semantic categories;
- (iii) alphabetical place of sub-entries; and
- (iv) chronological order of the categories of usage labels.

For example, given a word such as *put*, one proceeds as follows:

STEP 1

Find out a sentence, in which the basic meaning of a word is applied, e.g., *He put the book on the table*. This helps one to know where to start from in terms of the augmented valency and reduced valency, i.e. the increasing and decreasing complexity of the semantico-syntactic patterns shown below:

- (1) Ω \cong S
- (2) Ψ \cong SN
- (3) $K' \Psi$ \cong SCN
- (4) $\Psi \Phi$ \cong SF
- (5) $K \Psi$ \cong SC
- (6) $K \Psi \Phi$ \cong SCF
- (7) $K' K \Psi$ \cong SCC

They are related as follows in terms of complexity: (1) represents a single semantic-role pattern, (3) is derived from (2), (4) is derived from (2), (5) is derived from (2), (6) is derived from (4), and (7) is derived from (5). The derivation is indicative of the increasing and decreasing order of complexity of the semantico-syntactic patterns. The sentence *He put the books on the table* fits the

pattern $K \Psi \Phi \cong SCF$. The pattern is reducible to $\Psi \Phi \cong SF$ and finally to $\Omega \cong S$. Sentences that are likely to fit these patterns can be constructed using *put* as shown in Table VII, number (6) under STEP 2a. What should be noted is that according to the information in the Table VII, *put* does not fit in other patterns, i.e. (1)–(5) and (7) as indicated with dashes in the extreme right column of the Table.

STEP 2 (a)

Find the semantic-role with which the target word is used. This is done by constructing sentences based on the phrasal patterns in the ascending complexity of syntactico-semantic patterns, using that word. Examples are shown in Table VII.

Table VII: Construction of Sentences

Syntactico-semantic pattern	Sentence	Semantic-role
(1) $\Omega \cong S \rightarrow (N'')$, e.g., <i>The ball reddens.</i>	This pattern is not applicable to <i>put</i> .	—
(2) $\Psi \cong SN \rightarrow (N'' N'')$, e.g., <i>Ali is a doctor.</i> $(N'' A'')$, e.g., <i>The room is dirty.</i>	This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> .	— —
(3) $K' \Psi \cong SCN \rightarrow (N'' N'' N'')$, e.g., <i>They elected him president.</i> $(N'' N'' A)$, e.g., <i>They made the road wide.</i>	This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> .	— —
(4) $\Psi \Phi \cong SF \rightarrow (N'' Adv'')$, e.g., <i>The women were walking slowly.</i> $(N'' P'')$, e.g., <i>The children were sitting on a mat.</i>	This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> .	— —
(5) $K \Psi \cong SC \rightarrow (N'' N'')$, e.g., <i>He shattered the window.</i> $(N'' P'')$, e.g., <i>The visitors knocked at the door.</i> $(N'' S'[Cl_{that}, Cl_{\emptyset}, Cl_{to}, Cl_{wh-}, Cl_{-ing}, Cl_{-ed}])$. These stand for <i>that</i> clause, infinitive clause, zero infinitive clause, <i>wh</i> -clause, <i>-ing</i> clause and past participle	This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> .	— — —

clause respectively; for example, <i>He wanted me <u>to come</u></i> in the case of the infinitive clause.		
(6) $K \Psi \Phi \cong SCF \rightarrow$ (N" N" Adv") (N" N" P")	<i>He put <u>the box</u> <u>down</u>.</i> <i>The <u>pupil</u> put <u>the books</u> <u>on the table</u>.</i>	Ah Br Rl Ah Br ₁ Rr ₂
(7) $K' K \Psi \cong SCC \rightarrow$ (N" N" N"), e.g., <u>Jane</u> knitted <u>John</u> a sweater. (N" N" P"), e.g., <u>He</u> wrote a <u>letter with a pen</u> . (N" N" S'), <u>The florist</u> cut down <u>the trees to make the compound clean</u> .	This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> . This pattern is not applicable to <i>put</i> .	— —

STEP 2 b:

Using semantic category switching and semantic category variation, construct more sentences. From the structures shown above, *put* works in the semantico-syntactic pattern $K \Psi \Phi \cong SCF$. Under this pattern, the following sentences have been obtained: *He put the box down* | *The pupil put the books on the table*. A human being (h) occupies the subject position of the sentences (*he* and *the pupil*). A non-human being (r) occupies the direct object position (*the box* and *the books*). In addition, a non-human being (l and r) occupies the adverbial position (*down* and *on the table* respectively).

The categories in the sentences above can occupy new positions in the sentences; this is category switching. In the same vein, new categories can occupy some positions in the sentences; this is category variation. If category switching and variation occur, we wait to see if there is change of meaning of the target word. Category switching and variation results into the following positions of the categories for the sentences above:

Category switching: **Ah Br Rl**, Ah Bl Rr, Ar Bh Rl, Ar Bl Rh, Al Bh Rr, Al Br Rh
Category variation and switching: **Ah Br Rq**, Ah Bq Rr, Ar Bh Rq, Ar Bq Rh,
Aq Bh Rr

However, these patterns cannot yield valid sentences in terms of the word *put* because of category combinational constraints. However, in other cases they can (with category variation), for example, *The chef (h) put salt (r₁) in the food (r₂)* | *He (h) put the box (r) down (l)* | *She (h) put the bags (r) aside (l)*.

STEP 3a & b:

Formalise the sentences in STEP 2b, and examine the meaning of the word by

looking at each of the example sentences in terms of its formalisation. The sentences of *put* as constructed above are collected together and the meaning of *put* in each sentence is elicited.

He (h) put the box (r) down (l). put = to place something somewhere

The pupil (h) put the books (r₂) on the table (r₁). put = to leave something in a given place

The chef (h) put salt (r₁) in the food (r₂). put = to add something into another

She (h) put the bags (r) aside (l). put = to relocate an object from one place to another

The meanings of *put* for instance in Procter et al. (1995) which were elicited using the corpus (Procter 1995: viii), are comparable in texture to the ones above which are elicited via the semantico-syntactic approach. Procter et al. (1995) provides the following examples: *Put your clothes in the cupboard* | *Don't forget to put a stamp on the envelope* | *I need to put a new button on these trousers.* | *Come on, it's time to put these toys away.* The conceptual meanings of *put* in the examples above and similar others are given in a general form, *to move (something or someone) into the stated place, position or direction.* Compared to Summers (1987), the psychomotor meanings of *put* are lumped together as having one idea. They are recorded as follows: **put** /put/ v **put**, present participle **putting** [T] 1 [+ obj + adv/prep] to move, set, place, lay, or fix in, on or to a stated place: *Put the box on the table.* \ *Put the chair nearer the fire.* | *You put too much salt in this soup.* | *Put the toy in its box.* | *Put your hand over your mouth when you cough.* | *Put the newspaper down while I am talking to you.* However, according to the semantico-syntactic approach, the meanings would be split up depending on the entities the word *put* is describing; e.g. *to put a new button on the trousers* (Procter et al. 1995), which means 'to fix something on something'.

STEP 4:

Order the meanings according to ascending complexity of semantic-syntactic patterns (valency), ascending complexity of semantic categories, alphabetical place of sub-entries and chronological order of the categories of usage labels.

In terms of arrangement of meanings, the following criteria (the alphasyntactico-semantic mode) are applied:

- (i) Meanings will first be arranged according to the ascending order of semantic patterns, e.g., for *put* meanings were only in one pattern as:
 1. to place something somewhere: *He put the box down.* Ah Br RI
 2. to leave something in a given place: *The pupil put the books on the table.* Ah Br₁ Rr₂
 3. to add something into something: *The chef put salt in the food.* Ah Br₁ Rr₂
 4. to relocate an object from one place to another: *She put the bags aside.* Ah Br RI
- (ii) Within semantic patterns, meanings are arranged according to the ascend-

ing order of semantic categories, going by the categorial differences in the sentence formalisations. The semantic categories in their ascending order are: event (e), state (q), abstract (a); number (n), group (g), space (l), temporal object (t); concrete object (r), animate (b), human (h). (e), (q) and (a) represent nonphysical entities which obtain outside space and time, e.g., *write* as an act. (n), (g), (l), and (t) are nonphysical entities that obtain in space and time, e.g., *one* as a number. (r), (b), (h) represent physical objects that happen in space and time. Each of the objects in the three subgroups as presented presupposes the other. For example, (t, temporal object) presupposes (l, space) in the subgroup (n), (g), (l) and (t). Of the three subgroups, the second presupposes the first, while the third presupposes the second. Therefore, the ascending order of the semantic categories is based on the principle of existential presupposition; what presupposes the other comes last.

Thus, basing on the categorial differences of the sentence formalisations in (i) above, the arrangement of the formalisations according to the ascending order of the semantic categories is as follows:

- Ah Br Rl 1 add something into something: *The chef put salt in the to place something somewhere: Mary put the box down.*
- Ah Br Rl 2 to relocate an object from one place to another: *He put the bags aside.*
- Ah Br₁ Rr₂ 3 to leave something in a given place: *The pupil put the books on the table.*
- Ah Br₁ Rr₂ 4 to food.

(iii) Derived forms of a word, e.g. **written** (adj.) as in *English is a written language*, are arranged according to the alphabetical place of a sub-entry and entry. In the case of *put* there are no derived forms. Therefore, this particular criterion does not apply to the meanings of *put*.

(iv) Meanings that are a result of the same formalisation of sentences are arranged according to the chronological order of usage labels. The usage labels in their chronological order (determined according to what presupposes the other) are Origin (e.g. foreign origin); Time (e.g. archaic, obsolete, neologism); Status (e.g. standard, substandard); Variety (e.g. Scottish, Irish); Attitude (e.g. offensive, euphemism, disapproving); Social Group (e.g. the youth); Specialised (e.g. banking, physics, biology); Proprietary (i.e. a trademark meaning); and Medium (i.e. spoken form, written form).

Although, the following meanings in (ii) above have the same formalisation, i.e.,

- Ah Br Rl 1
- Ah Br Rl 2
- Ah Br₁ Rr₂ 3
- Ah Br₁ Rr₂ 4

none of them reflects the usage labels outlined above. Therefore, they will be arranged in the order in which they were elicited. However, this approach will be the source of inconsistency in the ordering of meanings, until the criterion for usage labels is further investigated.

Careful analysis of meanings in dictionaries, e.g. Summers et al. (2003), shows that meanings are categorised as meanings for simple word form (i.e. conceptual meanings), complex word form (*writer*), phrasal word form (*put up with*), compound word form (*air stream*), collocational word form (*put on hold*) and idiomatic word form (*kick the bucket*). The meanings of *put* as outlined above are only conceptual meanings. This implies that the rest of the kinds of meanings of *put* have not been arrived at using this method. However, one finds such meanings in dictionaries compiled using the corpus method. They are also the ones that are mostly arrived at in dictionaries compiled using the fieldwork methods, especially the collocational meanings. Such dictionaries do not present multiple conceptual meanings of a word due to the limitations of the fieldwork method used. The limitations include failure by informants to memorise the meanings of words to inform the interviewer, and high expenditure, that is, if a lexicographer has to use many more informants. Others are the cumbersome nature of the fieldwork method in terms of length of time involved and tediousness of the work.

5. The semantico-syntactic method and the alphasyntactico-semantic mode in Runyankore-Rukiga and Lugungu

Runyankore-Rukiga and Lugungu are among the Ugandan Bantu languages. In terms of writing, both languages fall in the category of less documented languages although Runyankore-Rukiga has quite a long history of writing than Lugungu. For instance, there is evidence of documentation of Runyankore-Rukiga that dates back to the late 1950s whereas the latest in Lugungu dates back to the 1990s. However, much as Runyankore-Rukiga was documented earlier than Lugungu, it does not have sufficient documentation that supports corpus-driven methods of compiling general-purpose dictionaries. The situation is aggravated by lack of money to establish corpus-driven dictionary projects. One of the interventions, particularly on the part of generating meaning of words and arranging them in a consistent order would be to use the semantico-syntactic method and the alphasyntactico-semantic mode. The examples below adopted from Oriikiriza (2011), show how the methods work in Runyankore-Rukiga and Lugungu.

5.1 Worked example for Runyankore-Rukiga using the word *okweta* (v.) "to call"

Using similar steps as in 4.3, we illustrate how the multiple meanings of the word *okweta* "to call" can be elicited and consistently arranged.

STEP 1:

Find out a sentence that expresses the basic meaning of a word.

- (i) The researcher in Oriikiriza (2011) asked an informant for a sentence in which the basic (or primary) meaning of *okweta* (v.) is expressed.
- (ii) In response, the following sentence was given as: *Irooko oyete Mugisha* "Go and call Mugisha".
- (iii) The informant was then asked for the meaning of *okweta* in this sentence.
- (iv) The informant replied that it means *okugambira omuntu ngu aije* "to tell someone to come over".
- (v) The researcher formalised the sentence as $Ah_1 Wh_2$.
- (vi) Based on the formalisation, the researcher identified its semantico-syntactic structure as $K\psi \cong SC (N'' S'')$.

STEP 2a:

Find out the semantic-role patterns with which the target word is used

- (i) With the pattern $K\psi \cong SC$, as in (iv) above, the researcher identified another immediate related pattern, i.e. $K' K\psi \cong SCC$ with its phrasal patterns: $(N'' N'' N'')$, $(N'' N'' P'')$ and $(N'' N'' S'')$.
- (ii) The researcher asked the informant for sentences that fit the phrasal patterns, and got the following responses:

$N'' N'' N''$ Nil.

$N'' N'' P''$ *Omwegyesa akeeta abeegi n'efurembe* "The teacher called the pupils using a whistle".

$N'' N'' S''$ Nil.

The researcher identified other distant patterns related to $K' K\psi \cong SCC$. These were $K\psi\Phi \cong SCF$ with the phrasal patterns $(N'' N'' Adv)$ and $(N'' N'' P)$, and $K' \psi \cong SCN$ with the phrasal patterns $(N'' N'' N)$ and $(N'' N'' A)$. The researcher asked the informant for sentences that fit the phrasal patterns above. The responses were given as shown in italicised forms below.

$K\psi\Phi \cong SCF$

$(N'' N'' Adv'')$ *Kato akeeta Kakuru aheeru* "Kato called Kakuru outside".

$(N'' N'' P'')$ *Nyineeka akeeta abantu aha bugyenyi* "The head of the household invited people to a party".

$K' \psi \cong SCN$

$(N'' N'' N'')$ *Abazaire bakeeta omwana eiziina* "The parents named the child".

$(N'' N'' A'')$ *Abahinguzi bakeeta omusinzi omushema* "The passers-by called the drunkard a fool".

- (iii) Eliciting the Meanings of *okweta* (v.) in the sentences above.

The informant was asked to give the meaning of *okweta* (v.) in each of the sentences above. The following meanings were recorded, and formalised.

- Sentence: *Irooko oyete Mugisha* (Ah₁ Wh₂) "Go and call your Mugisha".
Meaning of *okweta* (v.): to inform someone to come over.
Sentence: *Omwegyesa akeeta abeegi n'efurembe* (Ah₁ Wh₂ Rr) "The teacher called the pupils with a whistle".
Meaning of *okweta* (v.): *Okumanyisa abantu nari ekintu ngu kiije* "To make a signal to someone or something by means of something informing them to come over".
Sentence: *Kato akeeta Kakuru aheeru* (Ah₁ Wh₂ Rl) "Kato called Kakuru outside".
Meaning of *okweta* (v.): *okweta aheeru = okwiha omuntu omu bandi nari omu nju oine eki orikwenda kumugambira* "to call someone outside = to call someone from a group, or from a house when you have something you intend to tell them".
Sentence: *Nyineeka akeeta abantu aha bugyenyi* (Ah₁ Wh₂ Ra) "The head of the household invited people to a party".
Meaning of *okweta* (v.): *okutaayaayisa abantu* "to invite people to an occasion"
Sentence: *Abazaire bakeeta omwana eiziina* (Ah₁ Wh₂ Ra) "The parents named the child".
Meaning of *okweta* (v.): *okuha omuntu na munonga omwana, eiziina* "to give someone a name, especially children"
Sentence: *Abahinguzi bakeeta omusinzi omushema* (Ah₁ Wh₂ Rq) "The passers-by called the drunkard a fool".
Meaning of *okweta* (v.): *okugamba aha muntu oku ari* "to describe the behaviour of someone, or state in which someone is"

STEP 2b:

Using semantic category switching and semantic category variation, construct more sentences.

The researcher working with the informant established that semantic category variations are possible for the following patterns:

$K\psi\Phi \cong SCF$

(N" N" Adv") Nil.

(N" N" P") *Ebaruha neekweta aha bugyenyi* (Cr Wh Ra) "The letter is inviting you to a party".

Meaning of *okweta* (v.): *okumanyisa omuntu obutumwa bwe nookoresa ekindi kintu* "to inform someone by other means".

$K'\psi \cong SCN$

(N" N" N") *Ekya nitukyeta 'okuramy'a' omu Runyankore-Rukiga* (Ah Qa Re) "We call that 'okuramy'a' (worship) in Runyankore-Rukiga"

Meaning of *okweta* (v.): *okumanya ekintu omu nyetegyereza eyaawe* "to understand something in your own way"

Ekihandiiko tukyete 'Obuto bwangye' (Ah Qr Rq) Let us call the text "My Childhood"

Meaning of **okweta** (v.): *okuheereza ekintu eiziina ekitari muntu*, "to give a name to a non-human entity".

(N" N" A) Nil.

STEP 3a:

Formalise the sentences (i.e. express the sentences in terms of semantic-roles and semantic categories. (This was done simultaneously in STEPS 2a and 2b).

STEP 3b:

Examine the meaning of the word by looking at each of the examples in terms of its formalisation. (This was also done simultaneously in STEPS 2a and 2b).

STEP 4:

Order the meanings according to ascending complexity of semantic-syntactic patterns (i.e. valency), ascending complexity of semantic categories, alphabetical place of sub-entries and chronological order of the categories of usage labels.

Based on the criteria in (i)–(iii) above, which are the breakdown of the alphasyntactico-semantic order, the meanings of *okweta* (v.) were arranged as follows:

- (1) *okumanya ekintu omu nyetegyereza eyaawe*, "to understand something in your own way". *Ekyo nitukyeta 'okuramy'a' omu Runyankore-Rukiga* (Ah Qa Re) "We call that 'okuramy'a' (worship) in Runyankore-Rukiga"
- (2) *okuha omuntu namunonga omwana, eiziina*, "to give someone a name, especially children" Sentence: *Abazaire bakeeta omwana eiziina* (Ah₁ Wh₂ Ra) "The parents named the child".
- (3) *okuheereza ekintu eiziina ekitari muntu*, "to give a name to a non-human entity" *Ekihandiiko tukyete 'Obuto bwangye'* (Ah Qr Rq) "Let us call the text "My Childhood"
- (4) *okumanyisa omuntu obutumwa bwe nookoresa ekindi kintu*, "to inform someone by means of another". *Ebaruha neekweta aha bugyenyi* (Cr Wh Ra) "The letter is inviting you to a party".
- (5) *okugamba aha muntu oku ari*, "to describe the behaviour of someone, or manner in which someone is": *Abahinguzi bakeeta omusinzi omushema* (Ah₁ Wh₂ Rq) "The passers-by called the drunkard a fool".
- (6) *okutaayaayisa abantu*, "to invite people to an occasion": *Nyineeka akeeta abantu aha bugyenyi* (Ah₁ Wh₂ Ra) "The head of the household invited people to a party".
- (7) **okweta aheeru** = *okwiha omuntu omu bandi nari omu nju oine eki orikwenda kumugambira*, "to call someone outside = to call someone from a group of others, or from a house when you have something you intend to tell

them": *Kato akeeta Kakuru aheeru* (Ah₁ Wh₂ Rl) "Kato called Kakuru outside".

- (8) *okumanyisa abantu nari ekintu ngu kiije*, "To make a signal to someone or something by means of something, informing them to come over": *Omwegyesa akeeta abeege n'efurembe* Ah₁ Wh₂ Rr "The teacher called the pupils using a whistle".
- (9) *okugambira omuntu ngu aije* Ah₁ Wh₂: "to inform someone to come over" | *Irooko oyete Mugisha* "Go and call your Mugisha".

5.2 Example for Lugungu using the word *kudya* (v.) "to eat"

In the example below, adopted from Oriikiriza (2011), the process of eliciting meanings of the word *kudya* (v.) is in form of an interview between the researcher and the respondent. The elicitation was made more flexible than the one in the English example *put* and Runyankore-Rukiga *okweta* (v.). This is because the respondent was not a language specialist, but knowledgeable in the language. Even in the case of the Runyankore-Rukiga example, the respondent was not so much of a language specialist, and so the rigidity of the elicitation of meanings was lessened compared to the one seen in the English example, *put*. Against this background, the elicitation of meanings in Lugungu was done in an interview that mirrored STEPS 1–4 in the English and Runyankore-Rukiga examples. Below is the Lugungu elicitation and arrangement of meanings for the word *kudya* (v.)

STEP 1:

The Basic Meaning of *kudya* (v.) "to eat"

- (i) Interviewer: Construct a simple sentence that shows the core meaning of **kudya** (v.)
Respondent: *Badulu baadiiri bitooki byensei baabimala* "Men ate all the bananas and finished them"
- (ii) Interviewer: State the core meaning of **kudya** (v.) according to the example sentence above.
Respondent: to eat
- (iii) Interviewer: Formalisation of example sentence: *Badulu baadiiri bitooki byensei baabimala* "Men ate all the bananas and finished them". Ah Br
- (iv) Interviewer: Identification of semantico-syntactic pattern of the example sentence:
Kψ
- (v) Interviewer: Identification of the phrasal pattern of the example:
N" Adv"
N" N"
N" S'

(b) [Something or Somebody] [Verb], *Mwana adyambe* "The child eats"

- (ix) Interviewer: Formalisation of example sentence(s)
[Something or Somebody] [Verb], *Mwana adyambe* "The child eats", Wh
- (x) Interviewer: Give the meaning of **kudya** (v.) as implied in the example sentences.
Respondent:
[Something or Somebody] [Verb], *Mwana adyambe* "The child eats", Wh,
Here *adyambe* means habitual eating of food.

STEP 2b:

Using semantic category switching and semantic category variation, construct more sentences using the word *kudya* (v.).

Interviewer: In what different ways can you meaningfully vary the entities in the sentences above in STEP 1 and 2b? The sentences are repeated here below with guiding instructions:

- (i) For the sentence *Tito yaadiiri nduwa na nsu* (Titus ate posho with fish), construct other Lugungu sentences which mean [something (not a human-being)] [ate] [something] with [something].
- (ii) Based on the sentence, *Badulu baadiiri bitooki byensei baabimala* (Men ate all the bananas); construct other Lugungu sentences which express [something (not a human being)] [ate] [something], and indicate the meaning of *kudya* in that context.
- (iii) For the sentence *Mwana adyambe* (the child eats), construct other Lugungu sentences which show that [something (not human)] [Verb] using *kudya*, and indicate its meaning in that context.

Respondent:

Response to the first question: NIL. The respondent did not readily find the needed construction.

Response to first question: *Mbwene gyadiiri mbuli* "A dog ate meat." Contextual meaning of *kudya*:

Response to second question: The pattern [something (not human-being or animal)] [verb] [something] is not applicable in Lugungu.

However, the interviewer noticed that the respondent failed to arrive at the meaning "to each" which was expected for pattern above. Therefore, the interviewer asked the respondent as below.

Interviewer: What does *kudya* mean in each of the following context: *Mubiri gukundya* "The body itches me.", *Deeri twadiiri nyama* "Today we ate meat".

Respondent: *kudya* in *Mubiri gukundya* "The body itches me" denotes pain. *Kudya* in *Deeri twadiiri nyama* "Today we ate meat" denotes eating meat.

STEP 3a:

(Formalise the sentences (i.e. express the sentences in terms of semantic-roles and semantic categories. This was done simultaneously in STEPS 2a & b above.

STEP 3b: Examine the meaning of the word by looking at each of the examples in terms of its formalisation. This also was done simultaneously in STEP 2a & b above.

STEP 4:

Ordering the meanings of *kudya* (v.) according to ascending complexity of semantic-syntactic patterns (i.e. valency), ascending complexity of semantic categories, alphabetical place of sub-entries and chronological order of the categories of usage labels.

Interviewer: Collect and arrange the meanings together according to the alphasemantico-syntactic order.

Six meanings were obtained under the patterns: Ω , $K\psi$ and $K\psi\Phi$. The interviewer had to figure out the meanings because, from the responses, the meanings were not straightforward. The meanings are sequenced below:

- (1) to have the ability to eat: *Omwana adyambe* "The child eats", Wq
- (2) to take in something for a meal: *Deeru twadiiri nyama* "Today we ate meat", Ah Br
- (3) to consume something: *Badulu baadiiri bitooke byensi baabimala* "Men ate all the bananas and finished them" Ah Br
- (4) to devour: *Mbwene gyadiiri mbuli* "A dog ate a goat", Ab Br
- (5) to itch: *Mubiri gukundya* "The body itches me", Cr Bh
- (6) **kudya (ekintu) na (ekindi)** chewing and swallowing one type of food with another: *Tito yaadiiri nduwa na nsu* "Titus ate posho with fish". Ah Br₁ Mr₂

5.3 Observation on the Worked Examples

From the Runyankore-Rukiga worked example for the word *okweta* (v), nine meanings were generated and consistently arranged. For Lugungu, six meanings for the word *kudya* (v.) were generated and consistently arranged. Consequently, several more conceptual meanings are realised. As noted in section 4.1, these meanings result from sense and reference. If one were to add the usage and utterance meanings to these meanings, the set of all the meanings would be bigger.

6. Related works

Basic meaning and extended meanings as concepts are also discussed in the work of Pustejovsky and Rumshisky (2010) as well as Hanks (2013). Pustejovsky and Rumshisky (2010: 75) call the basic meaning an 'anchor sense' from

which extended meanings are derived. Thus, *in order to describe the sense extension mechanisms, we must assume, in a set of senses for a predicate, that one sense is an anchor, and the others are related through some transformations ... we are interested in identifying the formal mechanisms through which anchor sense(s) are extended to produce related senses.*

Examples of mechanisms for meaning extension as identified in Pustejovsky and Rumshisky (2010: 74, 78-79, 82-83) are described below (with illustrations and analyses from the same work):

- (1) entity characteristics: Different entities in the same argument structure, influence a distinctive meaning of the verb.

John killed the plant; CAUSE-TO-DIE (THING, ANIMATE).

Mary killed the conversation; TERMINATE (HUMAN, EVENT)

- (2) Scalar attributes: A predicate whose complement is a phrase expressing the level or scale will yield a metaphorical meaning. Whereas there is no metaphorical meaning expressed by the verb in *The plane arrived in New York on time*, there is one expressed by the verb in *The plane arrived at a cruising altitude of 30,000 feet*. Here, *arrive at* has a distinctive meaning according to Pustejovsky and Rumshisky (2010). It is a multiword verb compared to *arrive*.
- (3) Partitive connection: A predicate which expresses a partitive connection will carry a metaphorical meaning, depending on arguments with which it is used in a sentence. For instance, compared to *The boat anchored several miles away*, the word *anchor* in *A written constitution must be anchored in the idea of universal citizenship* has a metaphorical meaning.

This conceptualisation of how words gain meaning can be subsumed under sense and reference in the sections under the situation-role theory explained above.

The work in Hanks (2013: 105), exposes the reader to the corpus-driven methods of identifying meanings of words. Thus, *... different lexical sets in different syntactic roles can alter the meaning of the target word. This suggests that corpus analysis procedures for identifying lexical sets as clues for disambiguation are likely to be highly productive ...* (Hanks 2013: 105). As already noted in the earlier sections of this paper, corpus-driven methods of identifying and arranging meanings of words in less documented languages for purposes of dictionary compilation can be affected by limited written materials from which the corpus is generated. In such cases, it is feasible to use the mental screen basing on the semantic-syntactico-method and the alphasyntactico-semantic mode, to generate meanings of words and to arrange them consistently in dictionary form.

7. Conclusion

The semantico-syntactic method of elicitation of meanings of words and the alphasyntactico-semantic mode of arrangement of the meanings help lexicographers to compile good dictionaries in less documented languages. Good dic-

tionaries help to meet the meaning expectations of the prospective users of a given dictionary. Therefore, the semantico-syntactic method of elicitation of meanings of words and the alphasyntactico-semantic mode of arrangement of the meanings are approaches of dictionary-making whose end results meet such expectations. These are *inter alia*, conceptual meanings of words and consistent arrangement of the meanings.

References

- Brown, K. and J. Miller.** 1980. *Syntax: A Linguistic Introduction to Sentence Structure*. London/New York: Routledge.
- Hanks, P.** 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, Mass.: The MIT Press.
- Kiango, 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).
- Kiingi, K.B.** 2008. *The Semanticsyntactic Bridge I*. Available from: <<http://www.luganda.canadianwebs.com>> (accessed 10 September 2008).
- Kiingi, K.B.** 2009. *The Semanticsyntactic Bridge I*. Available from: <<http://www.luganda.canadianwebs.com>> (accessed 10 September 2009).
- Landau, S.I.** 1989. *Dictionaries: The Art and Craft of Lexicography*. New York: Cambridge University Press.
- Larson, Richard and Gabriel Segal.** 1995. *Knowledge of Meaning: An Introduction to Semantic Theory*. Original edition. Cambridge, Mass.: MIT Press.
- Le Veux, R.P.** 1917. *Premier essai de vocabulaire luganda-français d'après l'ordre étymologique*. Maison-Carrée, Algiers: Imprimerie des Missionnaires d'Afrique (Peres Blancs).
- Oriikiriza, C. (Ed.)**. 2007. *Kashoboorozi y'Orunyankore-Rukiga*. (Runyankore-Rukiga Dictionary.) Kampala: Fountain Publishers.
- Oriikiriza, C.** 2011. *Elicitation and Arrangement of Meanings of Words in the Lexicography of Bantu Languages*. Unpublished Doctoral Thesis. Makerere University, Kampala.
- Procter, P. et al (Eds.)**. 1995. *Cambridge International Dictionary of English* (ed. 1996, Reprint 1999). Cambridge: Cambridge University Press.
- Pustejovsky, J. and A. Rumshisky.** 2010. Mechanisms of Sense Extension in Verbs. De Schryver, G.-M. (Ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks: 67-88*. Kampala: Menha Publishers.
- Singh, R.A.** 1991. *An Introduction to Lexicography* (Reprint 1991). Mysore, India: Central Institute of Indian Languages (C.I.I.L).
- Soanes, C. (Ed.)**. 2001. *Oxford Dictionary of Current English*. Third edition. Oxford: Oxford University Press.
- Summers, D. (Ed.)**. 1987. *Longman Dictionary of Contemporary English*. New edition. Harlow: Longman.
- Summers, D. et al. (Eds.)**. 2003. *Longman Dictionary of Contemporary English: The Living Dictionary*. Fourth edition. Harlow: Pearson Education.
- Taylor, C.** 1959. *A Simplified Runyankore-Rukiga-English and English-Runyankore-Rukiga Dictionary*. Kampala: The Eagle Press, EALB.
- TUKI (Taasisi ya Uchunguzi wa Kiswahili)**. 1981. *Kamusi ya Kiswahili Sanifu*. Dar es Salaam: Oxford University Press and Institute of Kiswahili Research (IKR/TUKI), University of Dar es Salaam.

A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatisation — The *Groot Noord-Sotho Woordeboek* as a Case in Point*

D.J. Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (danie.prinsloo@up.ac.za)*

Abstract: This article gives a critical evaluation of the paradigm approach of the *Groot Noord-Sotho Woordeboek* to the lemmatisation of verbs and nouns derived from verbs. The verb stem *-roba* 'break' with its complicated system of derivations will be taken as a case in point. The paradigm presented for *-roba* will be evaluated in terms of structure, occurrence in Sepedi corpora and dictionaries, actual use by mother-tongue speakers, user-friendliness, contextualisation versus decontextualisation in relation to the cross-referencing system and space utilisation. Bringing together, and lexicographically treating all these forms for a single verb surely is a lexicographic achievement. The question, however, is to what extent such an approach is useful in respect of forms likely to be looked up by dictionary users, whether all of these forms actually exist, how user-friendly the approach and presentation is, if comment on semantics is sufficient and consistent and whether such a lumping approach actually saves space in contrast to entering derivations as main lemmas in a splitting approach.

Keywords: LEXICOGRAPHY, SEPEDI, LEXICOGRAPHIC TRADITIONS, LEMMATISATION STRATEGIES, THE VERB STEM *-ROBA*

Opsomming: 'n Kritiese evaluering van die paradigmenadering tot Sepedi-lemmatisering — Die *Groot Noord-Sotho Woordeboek* as voorbeeld.

Hierdie artikel gee 'n kritiese evaluering van die paradigmenadering tot die *Groot Noord-Sotho Woordeboek* tot die lemmatisering van werkwoorde en naamwoorde wat van werkwoorde afgelei is. Die werkwoordstam *-roba* 'breek' met sy komplekse sisteem van afleidings word as voorbeeld geneem. Die paradigma wat vir *-roba* aangebied word, sal in terme van struktuur, werklike gebruik deur moedertaalsprekers, voorkoms in Sepedikorpusse, gebruikersvriendelikheid, kontekstualisering versus dekontekstualisering ten opsigte van die kruisverwysingstelsel en ruimtebenutting geëvalueer word. Die byeenbring, en leksikografiese bewerking van al hierdie vorme vir 'n enkele

* This article was the basis of a shortened version presented as a paper at the Nineteenth Annual International Conference of the African Association for Lexicography (AFRILEX), which was hosted by the Research Unit for Language and Literature in the SA Context, North-West University, Potchefstroom Campus, Potchefstroom, South Africa, 1–3 July 2014.

werkwoord is sonder twyfel 'n leksikografiese prestasie. Die vraag is egter tot watter mate dit nuttig is ten opsigte van vorme wat waarskynlik deur woordeboekgebruikers opgesoek sal word, of al hierdie vorme werklik bestaan, hoe gebruikersvriendelik die benadering en aanbieding is, of semantiese kommentaar voldoende en konsekwent is, en of so 'n saamgevoegde benadering werklik ruimte bespaar in teenstelling met die aanbieding van afleidings as afsonderlike hooflemmas in 'n opgedeelde benadering.

Sleutelwoorde: LEKSIKOGRAFIE, SEPEDI, LEKSIKOGRAFIESE TRADISIES, LEMMATISERINGSTRATEGIEË, DIE WERKWOORDSTAM -ROBA

1. Introduction

The aim of this article is to give a critical evaluation of the paradigm approach to the lemmatisation of verbs, and nouns derived from verbs (deverbatives) in the *Groot Noord-Sotho Woordeboek/Comprehensive Northern Sotho Dictionary/Pukuntšú ya Sesotho sa Leboa* (Ziervogel and Mokgokong 1975). The complicated verb stem *-roba* 'break' will be taken as a case in point. The paradigm presented for *-roba* will be evaluated in terms of (a) structure, (b) real life use as reflected by occurrence in Sepedi (also referred to as Northern Sotho or Sesotho sa Leboa) corpora and dictionaries, as well as actual use by mother-tongue speakers, (c) user-friendliness of the paradigm approach in respect of lumping versus splitting, (d) contextualisation versus decontextualisation in relation to the cross-referencing system and (e) space utilisation. As a prerequisite, a brief theoretical background on lemmatisation approaches, traditions and strategies will be presented with special emphasis on the paradigm approach.

2. A brief theoretical background on lemmatisation approaches, traditions and strategies

Prinsloo (2009) distinguishes five aspects of importance for lemmatisation in African languages given in table 1. These aspects are discussed in more detail for nouns in Prinsloo and De Schryver (1999) and for verbs in Prinsloo (1994).

The Sepedi lexicographer has to deal with all of the aspects and subcategories in A to E in table 1. As far as A is concerned the *traditional* way to compile dictionaries, especially in the pre-corpus era was for the lexicographer to select lemmas on intuition/introspection. The advent of corpora enabled lexicographers to use *frequency* counts of words in a corpus as a major criterion for the inclusion or omission of lemmas. The *paradigm* approach could be described as an attempt to physically include all derivations of especially verbs in the dictionary. This is the approach in the *Groot Noord-Sotho Woordeboek* (GNSW) which will be outlined and evaluated in detail in this article. The *rule-orientated* approach stands in contrast to the paradigm approach in the sense that the urge is to reduce the number of lemmas presented for a specific paradigm to the absolute minimum. So, for example, only singular forms of nouns are lem-

matised and only stem forms (i.e. root + *-a*), without any extensions, of verbs are included as lemmas.

Table 2: Lemmatisation approaches, lexicographic traditions and lemmatisation strategies in Bantu languages

	A Lemmatisation approaches	B Orthography of the language	C Lexicographic traditions	D Lemmatisation strategies : verbs	E Lemmatisation strategies : nouns
1	Traditional	Disjunctive	Stem tradition	Strict stem	Strict stem
2	Rule-orientated	Conjunctive	Word tradition	Left-expanded stem	Left-expanded stem
3	Paradigm				Singular only
4	Frequency				Singular and plural
5					First and 3rd letter

Prinsloo (2009: 153)

The *orthography* of the language (B) plays an important role in the choice of the lexicographic tradition (C). A *disjunctively* written language such as Sepedi will e.g. write the phrase "I love you" as four orthographic words, i.e. *ke a go rata* and a conjunctively written language such as isiZulu as a single orthographic word, i.e. *ngiyakuthanda*. Both have exactly the same structure, i.e. subject concord + present tense marker + object concord + verb stem.

Disjunctively written languages such as Sepedi favour the *word tradition*, i.e. lemmatising nouns with their prefixes while the *stem* tradition is mostly chosen by lexicographers for conjunctively written languages. So, for example, *monna* 'man' will be lemmatised according to the word tradition on its full form under M while on its stem *-nna* under N in a stem dictionary such as GNSW.

3. The paradigm approach

In lexicography lumping versus splitting are mostly used in the literature in relation to the presentation of different senses of a word. In this article lumping versus splitting will be used in a grammatical sense i.e. grouping different derivations of a specific verbal stem under a single lemma or presenting each of the derivations as main lemmas. This brings the so-called paradigm approach following stem lemmatisation of GNSW in contrast with the traditional word lemmatisation approach of the *Pukuntšu* dictionaries (PUKU1 and PUKU2). GNSW lemmatises a verb under its stem form and all derivations of the verb including deverbatives will then be lumped together. The different forms will also be lemmatised *separately* as untreated lemmas with an implicit cross-refer-

ence to the main verb stem.

In the paradigm approach in GNSW the basic micro-architecture of an article is designed in terms of a modular layout aimed at bringing together all derivations of e.g. a verb stem. So, for example, the article of the lemma ROBA in GNSW consists of 32 modules distinguished on the basis of derived forms by suffixes and combinations of suffixes.

1. ROBA (-rôba, -rôbilê, -rôbja, -rôbilwê)
2. ROBAGANA (-rôbagana, -rôbagane, -rôbaganwa, -rôbaganwe)
3. ROBAGANELA (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe)
4. ROBAGANTŠHA (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê)
5. ROBAGANTŠHETŠA (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê)
6. ROBAGANTŠHETŠANA (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe)
7. ROBAGANYA (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê)
8. ROBAGANYETŠA (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê)
9. ROBAGANYETŠANA (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe)
10. ROBAKA (-rôbaka, -rôbakilê, -rôbakwa, -rôbakilwê)
11. ROBAKANA (-rôbakana, -rôbakane, -rôbakanwa, -rôbakanwe)
12. ROBAKANTŠHA (-rôbakantšha, -rôbakantšhitšê, -rôbakantšhwa, -rôbakantšhitšwê)
13. ROBAKANTŠHETŠA (-rôbakantšhêtsa, -rôbakantšhêditšê, -rôbakantšhêtswa, -rôbakantšhêditšwê)
14. ROBAKANTŠHETŠANA (-rôbakantšhêtsana, -rôbakantšhêtsane, -rôbakantšhêtsanwa, -rôbakantšhêtsanwe)
15. ROBAKANYA (-rôbakanya, -rôbakantšê, -rôbakanywa, -rôbakantšwê) ¹
16. ROBAKANYETŠA (-rôbakanyêtsa, -rôbakanyêditšê, -rôbakanyêtswa, -rôbakanyêditšwê)
17. ROBAKANYETŠANA (-rôbakanyêtsana, -rôbakanyêtsane, -rôbakanyêtsanwa, -rôbakanyêtsanwe)
18. ROBEGA (-robêga, -robêgilê)
19. ROBEGETŠA (-rôbêgêtsa, -rôbêgêditšê, -rôbêgêtswa, -rôbêgêditšwê)
20. ROBELA (-rôbêla, -rôbêtše, -rôbêlwa, -rôbêtšwe)
21. ROBELANA (-rôbêlana, -rôbêlane, -rôbêlanwa, -rôbêlanwe)
22. ROBESETŠA (-rôbêsêtsa, -rôbêsêditšê, -rôbêsêtswa, -rôbêsêditšwê)
23. ROBIŠA (-rôbiša, -rôbišitšê, -rôbišwa, -rôbišitšwê)
24. ROBIŠANA (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe)
25. ROBOKA (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê)
26. ROBOKANA (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe)
27. ROBOKANTŠHA (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê)
28. ROBOKANTŠHETŠA (-rôbôkantšhêtsa, -rôbôkantšhêditšê, -rôbôkantšhêtswa, -rôbôkantšhêditšwê)
29. ROBOKANTŠHETŠANA (-rôbôkantšhêtsana, -rôbôkantšhêtsane, -rôbôkantšhêtsanwa, -rôbôkantšhêtsanwe)
30. ROBOKANYA (-rôbôkanya, -rôbôkantšê, -rôbôkanywa, -rôbôkantšwê)
31. ROBOKANYETŠA (-rôbôkanyêtsa, -rôbôkanyêditšê, -rôbôkanyêtswa, -rôbôkanyêditšwê)
32. ROBOKANYETŠANA (-rôbôkanyêtsana, -rôbôkanyêtsane, -rôbôkanyêtsanwa, -rôbôkanyêtsanwe)

The full article of ROBA in GNSW is given in the appendix.

In module 1 the lemma is the basic stem (root *-rob-* plus the terminative *-a*) without any suffixes. The stem is repeated followed by the perfect, passive and perfect plus passive forms. Prinsloo and De Schryver (1999) refer to the latter three as "standard modifications". Modules 2-32 give the root plus a suffix or combination of suffixes with the standard modifications. For example, ROBELANA in Module 21 consists of the root plus applicative suffix (*-el-*) plus reciprocal

(-an-) plus the verbal ending followed by the perfect form *-rôbêlane*, passive *-rôbêlanwa*, and perfect plus passive *-rôbêlanwe*.

The module layout includes comments on form and on semantics, mainly giving translation equivalents in Afrikaans and English as well as examples of usage and deverbatives.

Consider the treatment of ROBAGANA in Module 2:

2. ROBAGANA (-rôbagana, -rôbagane, -rôbaganwa, -rôbaganwe) ass.; in stukke gebreek w. // b. broken into pieces; morobagani, ba- (morôbagani) pers. dev.; morobagano, me- (morôbaganô) man. dev.

Treatment in this module includes four nouns which are derived from *-robagana* i.e. *morobagani*, *barobagani*, *morobagano* and *merobagano*.

The entire article of ROBA consists of 265 nominal and verbal forms of *roba*:

barobagani, barobagantšhetšani, barobagantšhetši, barobagantšhetšo, barobagantšhi, barobaganyetšani, barobaganyetši, barobaganyi, barobakantšhetšani, barobakantšhetši, barobakantšhi, barobakanyetšani, barobakanyetši, barobakanyi, barobaki, barobedi, barobegetši, barobegi, barobelani, barobeseditši, barobi, barobišani, barobiši, barobja, barobokantšhetšani, barobokantšhetši, barobokantšhi, barobokanyetšani, barobokanyetši, barobokanyi, baroboki, barokantšhetšani, borobaganelo, diroba, dirobaganedi, dirobakani, dirobakantšhwa, dirobakanywa, dirobja, dirobokani, dirobokantšhwa, dirobokantšwa, dirobokanywa, dithobego, ithoba, ithobela, merobaganelo, merobagano, merobagantšhetšano, merobagantšho, merobaganyetšano, merobaganyetšo, merobaganyo, merobakano, merobakantšhetšano, merobakantšhetšo, merobakantšho, merobakanyetšano, merobakanyetšo, merobakanyo, merobako, merobegetšo, merobego, merobelano, merobelo, merobeseditšo, merobišano, merobišo, merobokano, merobokantšhetšano, merobokantšhetšo, merobokantšho, merobokanyetšano, merobokanyetšo, merobokanyo, meroboko, morobaganelo, morobagani, morobagano, morobagantšhetšano, morobagantšhetši, morobagantšhetšo, morobagantšhi, morobagantšho, morobaganyetšano, morobaganyetši, morobaganyetšo, morobaganyi, morobaganyo, morobakano, morobakantšhetšano, morobakantšhetši, morobakantšhetšo, morobakantšhi, morobakantšho, morobakanyetšano, morobakanyetši, morobakanyetšo, morobakanyi, morobakanyo, morobaki, morobako, morobedi, morobegetši, morobegetšo, morobegi, morobego, merobelano, merobelo, morobesetši, morobesetšo, morobi, morobišano, morobiši, morobišo, morobja, morobokano, morobokantšhetšano, morobokantšhetši, morobokantšhetšo, morobokantšhi, morobokantšho, morobokanyetšano, morobokanyetši, morobokanyetšo, morobokanyi, morobokanyo, moroboki, moroboko, roba, robagana, robagane, robaganela, robaganelwa, robaganetše, robaganetšwe, robagantše, robagantšha, robagantšheditše, robagantšheditšwe, robagantšhetša, robagantšhetšana, robagantšhetšane, robagantšhetšanwa, robagantšhetšanwe, robagantšhetšwa, robagantšhitšwe, robagantšhitše, robagantšhwa, robagantšwe, robaganwa, robaganwe, robaganya, robaganyeditše, robaganyeditšwe, robaganyetša, robaganyetšana, robaganyetšanwa, robaganyetšanwe, robaganyetšwa, robaganywa, robagenela, robaka, robakana, robakane, robakantše, robakantšha, robakantšheditše, robakantšheditšwe, robakantšhetša, robakantšhetšana, robakantšhetšane, robakantšhetšanwa, robakantšhetšanwe, robakantšhetšwa, robakantšhitše, robakantšhitšwe, robakantšhwa, robakantšwe, robakanwa, robakanwe, robakanya, robakanyeditše, robakanyedi-

tšwe, robakanyetša, robakanyetšana, robakanyetšane, robakanyetšanwa, robakanyetšanwe, robakanyetšwa, robakanywa, robakile, robakilwe, robakwa, robega, robegeditše, robegeditšwe, robegetša, robegetšwa, robegile, robela, robelana, robelane, robelanwa, robelanwe, robelwa, robeseditše, robeseditšwe, robesetša, robesetšwa, robetše, robetšwe, robile, robilwe, robiša, robišana, robišane, robišanwa, robišanwe, robišitše, robišitšwe, robišwa, robja, roboka, robokana, robokane, robokantše, robokantšha, robokantšheditše, robokantšheditšwe, robokantšhetša, robokantšhetšana, robokantšhetšane, robokantšhetšanwa, robokantšhetšanwe, robokantšhetšwa, robokantšhitše, robokantšhitšwe, robokantšhwa, robokantšwe, robokanwa, robokanwe, robokanya, robokanyeditše, robokanyeditšwe, robokanyetša, robokanyetšana, robokanyetšane, robokanyetšanwa, robokanyetšanwe, robokanyetšwa, robokanywa, robokile, robokilwe, robokwa, seroba, serobagamedi, serobakantšhwa, serobakanywa, serobja, serobokantšhwa, serobokantšwa, serobokanywa, thobego

Certain omissions in the paradigm completion were however noticed:

1. Module 6: *morobagantšhetšani* (only the plural *barobagantšhetšani* is given)
2. Module 9: *morobaganyetšani* (only the plural *barobaganyetšani* is given)
3. Module 11: *serobakani* to be added
4. Module 14: *morobakantšhetšani* to be added
5. Module 17: *morobakanyetšani* to be added
6. Module 21: *morobelani* to be added
7. Module 24: *morobišani* to be added
8. Module 26: *serobokani* to be added
9. Module 29: *morobokantšhetšani* to be added
10. Module 31: *morobokanyetšani* to be added

Accounting for, and to lexicographically treat all these forms for a single verb surely is a lexicographic achievement and succeeds in bringing together all derived forms of *-roba* in a single article. The question, however, is to what extent is it useful to the dictionary user in respect of (a) how many of these forms are likely to be looked-up by the users, (b) do all of these forms actually exist, (c) how user friendly is the approach/presentation, (d) is comment on semantics sufficient and consistent, (e) does such a lumping approach really save space in contrast to giving derivations as main lemmas? These issues will be briefly addressed in the following paragraphs.

4. Usage versus presumed usage and existence of words in the language

Ziervogel (1965) says that the basic meaning of a word lies in its root, e.g. for *-roba* in *-ROB-* and by adding a series of pre- and suffixes the root can obtain a variety of senses/meanings which have to do with the basic meaning.

Hoewel die wortel selde 'n selfstandige gebruik in die taal het, dui hy wel altyd 'n begrip aan ... Deur die toevoeging van 'n reeks voor- en/of agtervoegsels kan die wortel 'n verskeidenheid begrippe kry wat met die basiese betekenis te doen het. Die voor- en agtervoegsel het wel 'n betekenisinhoud maar nie noodwendig

'n ekwivalente betekenis in Afrikaans nie. (Ziervogel 1965: 47)

Thus for *-roba* it means that *-ROB-* is the ideal point of departure for building a paradigm of derivations by means of affixes in order to reflect the variety of different meanings. In Ziervogel's view this also means that stem lemmatisation is the ideal lemmatisation strategy, e.g. for a systematised representation of word formation.

Ziervogel (1965: 45) claims that: Entries must be arranged under their stems with cross-references where necessary. This method is scientifically sound. A systematised survey of word formation in the languages is given; it shows word and lexical relationship and prevents repetition. (Van Wyk 1995: 85)

Van Wyk's severe and detailed criticism of the GNSW's approach is focused on the deficiencies of employing a stem lemmatisation strategy instead of a word lemmatisation strategy for a disjunctively written language such as Sepedi. He rejects Ziervogel's claims that stem lemmatisation is scientifically more sound than word lemmatisation, or that it prevents repetition. Of special importance to this article is Van Wyk's statement that it is the task of a grammar book and not a dictionary to give a systematic survey of word formation.

In this article the focus is on presumed aspects of user-unfriendliness in relation to problematic aspects of the presentation and especially the selection of lemmas.

As for the selection of lemmas Ziervogel acknowledges inclusion versus omission of lemmas as being important and problematic and suggests that the written language should be the point of departure for an effort to include all written forms.

Die probleem van wat in 'n woordeboek opgeneem moet word, is nogal van belang ... Ek glo 'n mens moet begin by die geskrewe taal en alle geskrewe woorde opneem. (Ziervogel 1965: 50)

He continues that for a comprehensive dictionary it is important to document the derivations but that the question is to what extent reduplications (repetition of a word with added affixes) and reflexives (i.e. do something to oneself) should be included.

Vir 'n redelik volledige woordeboek is dit tog van belang om aan te teken watter afleidings gemaak kan word. Die vraag is natuurlik vir hoever afleidings soos reduplikasie en refleksiewe opgeneem moet word. (Ziervogel 1965: 52)

This brings us to the core of the issue, i.e. what is the duty of the lexicographer in terms of what to include and what to omit from the dictionary.

Gouws and Prinsloo (2005: 86) state that in a general dictionary (with a text reception function) "the user should be able to find the words encountered in the day to day general language usage ...". The lexicographer should include a selection from the lexical stock of the language. It should not be limited to

words found in written texts but also include words from the spoken language. The question, however is whether the lexicographer should invent words in the sense of e.g. derivations that are theoretically possible, i.e., that could possibly be derived in terms of the grammatical rules of the language? In the case of *-roba* one would have to ask whether all of the derivations given in GNSW are really in use in the language and what the likelihood would be for them to actually be looked up by the target users of the dictionary.

In order to determine the likelihood of the different derivations of *-roba* to be looked up as well as their actual use in the language, (a) their actual occurrence in the Pretoria Sepedi Corpus (PSC) was determined, (b) the treatment of *-roba* in Sepedi dictionaries was studied and (c) two mother tongue speakers of Sepedi were requested to indicate which of these forms they know.

As for the PSC, only 35 (13%) of the 265 words occurred in the corpus, frequencies given in brackets.

robetše (642)*, roba (229), robega (119), robile (81), robegile (63), robja (60), seroba (40), diroba (28), robetšwe (25)*, robaganya (24), robela (14), thobego (13), robilwe (13), ithoba (6), roboka (5), robagana (3), robagantšwe (3), robesetša (3), robaka (2), robelwa (2), robokanya (2), robokile (2), robiša (2), robaganywa (2), robaganyeditše (1), robeseditšwe (1), robagantšha (1), robaganyetša (1), robagantše (1), robagane (1), ithobela (1), ithobela (1), robakanya (1), robagantšha (1), robokwa (1)

*Frequency count includes homonyms: *thoko* 'side', *robetše* 'sleep, slept', *robetšwe* 'was slept'

Only ten derivations of *-roba* are lemmatised and treated in seven randomly selected Sepedi dictionaries as given in table 2.

Table 2: Derivations of *-roba* in Sepedi dictionaries

POP=*Popular Northern Sotho Dictionary*

NEND=*New English–Northern Sotho Dictionary*

PUKU1=*Pukuntšu Dictionary (1983)*

SLEPD=*Sesotho sa Leboa/English Pukuntšu Dictionary*

ONSD=*Oxford Bilingual School Dictionary: Northern Sotho and English*

PUKU2=*Pukuntšu Dictionary (1989)*

PTLH=*Pukuntšutlhaloši ya Sesotho sa Leboa*

	POP	NEND	PUKU1	SLEPD	ONSD	PUKU2	PTLH
roba	√	√	√	√	√	√	√
robega	√	√	√	√	√	√	√
robaganya	√	√	√			√	
robja	√	√	√	√			
robile	√			√	√		
robakanya			√			√	
robaka		√					
robela		√					
robegile	√						
robagantše	√						

From the 265 forms of *-roba* the mother tongue speakers only identified 54 (20%) and 77 (29%) respectively as known to them. Forms which they both know are indicated in bold.

barobagani	merobego	robaganela	robelana
barobagantšhetšani	merobelano	robaganelwa	robelane
barobagantšhi	merobelo	robaganetše	robelanwe
barobaganyi	merobišano	robagantše	robelwa
barobakanyi	merobišo	robagantšha	robesešša
barobaki	merobokantšho	robagantšheditše	robetše
barobegi	merobokanyo	robagantšheditšwe	robetšwe
barobelani	meroboko	robagantšhetša	robile
barobi	morobagano	robagantšhetšana	robilwe
barobišani	morobagantšhi	robagantšhetšane	robišana
barobiši	morobaganyi	robagantšhitše	robišana
barobja	morobakantšhi	robagantšwe	robišane
barobokantšhetšani	morobakantšho	robaganya	robišitše
barobokantšhi	morobakanyi	robaka	robišitšwe
barobokanyi	morobegetši	robakana	robišwa
borobaganelo	morobegi	robakane	robja
diroba	morobego	robakantše	robokantše
dirobakantšhwa	morobelano	robakantšha	robokantšha
dirobja	merobelo	robakantšhitše	robokanya
dirobokanywa	morobeseššo	robakantšhitšwe	seroba
dithobego	morobi	robakantšwe	serobagamedi
ithoba	merobišano	robakanya	serobakantšhwa
ithobela	merobiši	robakile	serobakanywa
merobagantšho	merobišo	robakwa	serobja
merobaganyo	roba	robega	serobokanywa
merobakano	robagana	robegile	thobego
merobegetšo	robagane	robela	

The very limited number of derivations of *-roba* that (a) occur in the Sepedi Corpus, (b) are lemmatised and treated in Sepedi dictionaries and (c) were recognised by the mother-tongue speakers, suggest that the actual extent of use of derivations of *-roba* is much smaller than suggested by GNSW.

The question is thus whether most of these words given by GNSW for *-roba* actually *exist* in the language or whether the compilers mainly focused on categorically completing morphological/grammatical paradigms? Does the task of the lexicographer go beyond the lemmatisation and treatment of words actually used in a language to those that can *potentially exist* because they are e.g. possible through morphological reduplication rules or might come into use as new inventions? As for the latter it is true that one could never claim that a specific reduplication/combination will never come into use. For example, the author once objected to the inclusion of *deurgans* as a noun 'door goose' in a spelling checker lexicon for Afrikaans only to find that goose door stops exist in English, cf. images at http://vintagepatterns.wikia.com/wiki/Patch_Press_379.

It is therefore possible that *deurgans* can become a common word in Afrikaans. As a second example, it would have been unlikely a few years ago to include a nominal *like(s)* as a lemma in a dictionary but today it is commonly used on websites, e.g. 34 likes.

The English language is notoriously fast in adapting to the changing world. New words enter English from every area of life where they represent and describe the changes and developments that take place from day to day. Here are some words and expressions that have been coined in recent years. Some can be found in official dictionaries; others may never make their way there, but new words will continue to appear as the English language adapts to innovations and trends:

<http://www.learn-english-today.com/new-words/new-words-in-english.html>

Consider the following examples stated: *breadcrumbing* (a navigation technique which helps users by displaying a list of links to the pages they have visited when exploring a website), *copyleft* (opposite of copyright ... allows freedom of use for all), *crowdfunding* (raising money for a project by getting a large number of people to make a small financial contribution), *cyberbully* (a person who uses the Internet to harm another person), *textspeak* (language used in text messages), etc. <http://www.learn-english-today.com/new-words/new-words-in-english.html>

Be that as it may, it is not the task of the lexicographer to provide for possible future use/existence of words in a dictionary. Lemma selection should not be influenced by words that the lexicographer would like to see as part or to become part of the language. In terms of Wells (1973), Hartmann (1983) and Gove (1961) the duty of the lexicographer is to record language and to include words which are actually used by the speakers of the language in the dictionary.

The responsibility of a dictionary is to record the language, not set its style ... The only area in which the truth may be found is actual usage. In fine, the function of a dictionary is to reflect the facts of usage as they exist. A dictionary neither permits nor prevents. (Wells 1973: 84)

Lexemes become entries in a dictionary only when they are socialised, that is when they are used by a sufficient number of speakers. (Hartmann 1983: 71)

The basic aim is nothing less than coverage of the current vocabulary of standard spoken and written English. (Gove 1961: 4a)

The lexicographer's attention should be limited to the treatment of existing words in the lexicon especially given the fact that it is hardly possible to cover the existing words, even in a comprehensive multivolume dictionary. Currently available corpora which reflect actual use of words and indicating their frequency of occurrence are the ideal sources to guide the lexicographer in the selection of lemmas.

5. User-friendliness of the paradigm approach

Gouws and Prinsloo (2005: 39) emphasize the importance of the user-perspective:

The user-perspective, so prevalent in modern-day metalexigraphy, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. Good dictionaries are products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are.

Bothma and Prinsloo (2013) emphasize that the user may not want to read or browse through a long article with much irrelevant information in terms of his/her specific information need at a given time. In most cases (s)he only requires the information needed to solve the current information need. The lexicographer should therefore guard against excessive offering of information and rather guide the user more directly to the required information. Haas' remark of five decades ago still holds true:

A good dictionary is one in which you can find the information you are looking for — preferably in the very first place you look. (Haas 1962: 48)

Consider in this regard Prinsloo et al. (2011) where users are guided through decision trees directly to the required information.

An approach to lump information together in long dictionary articles as for ROBA in GNSW runs against the desire to quickly and directly find the information that the user is looking for at a given time.

5.1 Evaluation of GNSW in terms of user-friendliness

The GNSW is generally regarded as user-unfriendly. Prinsloo and De Schryver (1999: 258) state that the user-perspective was not seriously considered in the compilation. Their main criticism in this regard is against the use of phonemic sorting on lemmas, and stem lemmatisation. As for the sorting order the compilers of GNSW deviate from an ordinary alphabetical sorting of the entries and utilize a phonemic one, namely: A, B, BJ, D, E, F, FS, FŠ, G, H, HL, I, J, K, KG, KH, L, M, N, NG, NX, NY, O, P, PH, etc., because this is in their opinion 'more scientific'. To the user it is nothing more than sheer frustration to eventually find, for example, a word commencing on *bj* alphabetically after *bu* in the dictionary, (Prinsloo and De Schryver 1999: 261).

The layout of the complex article of *-roba* as given in the appendix is user-unfriendly in many ways. First it is very long. Secondly, although the derivations are alphabetically ordered as sublemmas, they are presented in a run-on layout which makes it difficult to detect them as the starting point for most of the 32 modules. Thirdly the use of capital letters to mark them is compromised by the use of the same convention to indicate the derivation from which a specific sublemma was derived. Consider Modules 3-9 in this regard.

3. ROBAGANELA (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe)
4. ROBAGANTŠHA (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê)
5. ROBAGANTŠHETŠA (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê)
6. ROBAGANTŠHETŠANA (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe)
7. ROBAGANYA (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê)
8. ROBAGANYETŠA (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê)
9. ROBAGANYETŠANA (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe)

For each of the starting points for Modules 3-9, it is indicated by means of the symbol "<" that they are derived from *robagana*, i.e. < ROBAGANYA:

(morôbaganô) man. dev.; **ROBAGANELA** (-rôbagenêla, -rôbaganêtše, -rôbaganêlwa, -rôbaganêtšwe) appl. < **ROBAGANA**; **borobaganelo** (borôbaganêlô) lo. dev.; **morobaganelo**, **me-** (morôbaganêlô) man. dev.; **serobagenedi**, **di-** (serôbaganêdi) impers. dev.; **ROBAGANTŠHA** (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê) caus. < **ROBAGANA**; in stukke breek // break into pieces; **morobagantšhi**, **ba-** (morôbaganantšhi) pers. dev.; **morobagantšho**, **me-** (morôbaganantšhò) man. dev.; **ROBAGANTŠHETŠA** (-rôbagantšhêtsa, -rôbagantšhêditšê, -rôbagantšhêtswa, -rôbagantšhêditšwê) caus. appl. < **ROBAGANA**; **morobagantšhetši**, **ba-** (morôbaganantšhetši) pers. dev.; **morobagantšhetšo**, **me-** (morôbaganantšhetšò) man. dev.; **ROBAGANTŠHETŠANA** (-rôbagantšhêtsana, -rôbagantšhêtsane, -rôbagantšhêtsanwa, -rôbagantšhêtsanwe) caus. appl. rec. < **ROBAGANA**; **barobagantšhetšani** (barôbaganantšhetšani) pers. dev.; **morobagantšhetšano**, **me-** (morôbaganantšhetšanò) man. dev.; **ROBAGANYA** (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê) caus. < **ROBAGANA**; in stukke breek // break into pieces; **morobaganyi**, **ba-** (morôbagananyi) pers. dev.; **morobaganyo**, **me-** (morôbagananyò) man. dev.; **ROBAGANYETŠA** (-rôbaganyêtsa, -rôbaganyêditšê, -rôbaganyêtswa, -rôbaganyêditšwê) caus. appl. < **ROBAGANA**; **morobaganyetši**, **ba-** (morôbagananyetši) pers. dev.; **morobaganyetšo**, **me-** (morôbagananyetšò) man. dev.; **ROBAGANYETŠANA** (-rôbaganyêtsana, -rôbaganyêtsane, -rôbaganyêtsanwa, -rôbaganyêtsanwe) caus. appl. rec. < **ROBAGANA**;

This way of indicating the source of derivation in a run-on layout thus obscures the capitalised starting point of the modules making it more difficult for the user to find the sublemma. Starting each of the 23 modules on a new line would have substantially increased user-friendliness of the layout.

5.2 Inadequate comment on semantics

The predicament of the user however does not end with the difficulty of locating the specific derivation for which (s)he wants to find the meaning. In most cases (s)he will find the specific sublemma with its presumed standard modifications neatly spelled out but without any comments on semantics. The use of actual comment on semantics in the article of ROBA is very limited, especially in relation to the length of the article. So, for example, no comment on semantics is given for the entire stretch of modules 24-32, i.e. ROBIŠANA to ROBOKANYE-TŠANA:

me- (morôbišô) man. dev.; *ROBIŠANA* (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe) caus. rec.; *barobišani* (barôbišani) pers. dev.; *morobišano*, me- (morôbišanô) man. dev.; *ROBOKA* (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê) it., v. *ROBAKA*; *moroboki*, ba- (morôbôki) pers. dev.; *moroboko*, me- (morôbôkô) man. dev.; *ROBOKANA* (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe) rec. < *ROBOKA*; v. *ROBAKANA*; *dirobokani* (dirôbôkani) pers. dev.; *morobokano*, me- (morôbôkanô) man. dev.; *ROBOKANTŠHA* (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < *ROBOKANA*; (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < *ROBOKANA*; v. *ROBAKANTŠHA*; *morobokantšhi*, ba- (morôbôkantšhi) pers. dev.; *morobokantšho*, me- (morôbôkantšhó) man. dev.; *serobokantšhwá*, di- (serôbôkantšhwa) impers. pass. dev.; *ROBOKANTŠHETŠA* (-rôbôkantšhêtša, -rôbôkantšhêditšê, -rôbôkantšhêtšwa, -rôbôkantšhêditšwê) caus. appl. < *ROBOKANA*; *morobokantšhetši*, ba- (morôbôkantšhêtši) pers. dev.; *morobokantšhetšo*, me- (morôbôkantšhêtšó) man. dev.; *ROBOKANTŠHETŠANA* (-rôbôkantšhêtšana, -rôbôkantšhêtšane, -rôbôkantšhêtšanwa, -rôbôkantšhêtšanwe) appl. rec. < *ROBOKANTŠHA*; *barobokantšhetšani* (barobokantšhêtšani) pers. dev.; *morobokantšhetšano*, me- (morôbôkantšhêtšanô) man. dev.; *ROBOKA-*

NYA (-rôbôkanya, -rôbôkantšê, -rôbôkanywa, -rôbôkantšwê) caus. < **ROBOKANA**; v. **ROBAKANYA**; morobokanyi, ba- (morôbôkanyi) pers. dev.; morobokanyo, me- (morôbôkanyô) man. dev.; serobokanywá, di- (serôbôkanywa) impers. pass. dev.; **ROBOKANYETŠA** (-rôbôkanyêtšá, -rôbôkanyêditšê, -rôbôkanyêtšwa, -rôbôkanyêditšwê) appl. < **ROBAKANYA**; morobokanyetši, ba- (morôbôkanyêtši) pers. dev.; morobokanyetšo, me- (morôbôkanyêtšô) man. dev.; **ROBOKANYETŠANA** (-rôbôkanyêtšana, -rôbôkanyêtšane, -rôbôkanyêtšanwa, -rôbôkanyêtšanwe) appl. rec. < **ROBOKANYA**; barobokanyetšani (barôbôkanyêtšani) pers. dev.; morobokanyetšano, me- (morôbôkanyêtšanô) man. dev.

This reflects a serious imbalance between comment on form versus comment on semantics which is detrimental to the main reason for looking up words in a dictionary, i.e. to find its meaning. Gouws and Prinsloo (2005: 48) refer to the "main assignment" of linguistic dictionaries "i.e. to give an explanation of the meaning of the lemma in monolingual dictionaries and to provide target language translation equivalents for a source language lemma in bilingual and multilingual dictionaries". It could be argued that the compilers of GNSW were so obsessed to include all possible derived forms that comment on semantics was neglected. Prinsloo and De Schryver (1999: 261) call it an 'enter-them-all-syndrome'.

In the article of *phêfa* the compilers apparently concentrated so hard on completing the modular paradigms that they 'forgot' to give any translation equivalents in Afrikaans and English for the entire article.

PHÉFA (-phêfa, -phêfilê, -phêfša, -phêfilwê) (Pb.) v. **PHEMA**; mophêfi, ba- (mophêfi) pers. dev.; phêfo, (n-)/di- (phêfô) man. dev.; v. **phemo**; **PHÉFANA** (-phêfana, -phêfane, -phêfanwa, -phêfanwe) rec.; v. **phemana**; baphêfani (baphêfani) pers. dev.; phêfáno, (n-)/di- (phêfanô) man. dev.; **PHÉFÉGA** (-phêfêga, -phêfêgilê) neutr.; v. **phemega**; **PHÉFÉLA** (-phêfêla, -phêfêtše, -phêfêlwa, -phêfêtšwe) appl.; mophêfédi, ba- (mophêfédi) pers. dev.; phêfélo, (n-)/di- (phêfélô) man. dev.; **PHÉFÉLANA** (-phêfélana, -phêfélane, -phêfélanwa, -phêfélanwe) appl. rec.; baphêfélani (baphêfélani) pers. dev.; phêfélano, (n-)/di- (phêfélano) man. dev.; **PHÉFÍŠA** (-phêfiša, -phêfišitšê, -phêfišwa, -phêfišitšwê) caus.; mophêfiši, ba- (mophêfiši) pers. dev.; phêfišo, (n-)/di- (phêfišô) man.

dev.; PHÉFÍŠANA (-phêfišana, -phêfišane,
-phêfišanwa, -phêfišanwe) caus. rec.; ba-
phêfišani (baphêfišani) pers. dev.; phêfišano,
(n-)/di- (phêfišanô) man. dev.

5.3 Efficiency of the medio-structure

The lumping approach in GNSW also reduces the effectiveness of the medio-structure (system of cross-referencing) which is crucial in a lumping approach i.e. to guide the users from a reference position outside the article where the derivation was lemmatised in the alphabetical stretch, to the reference address inside the main article where the derivation in question is treated. Gouws and Prinsloo (2005: 181) state that one of the important functions of the medio-structure of a dictionary is to combat the decontextualisation brought about by alphabetical ordering. In a simplified way one could say that alphabetical ordering of lemmas in a dictionary has the detrimental effect of decontextualizing words that belong together. By way of comparison, words indicating *fruit* such as *apple*, *pear*, *banana* and *orange* belong together but are scattered over the dictionary as they belong to different alphabetical stretches. Dictionaries consequently attempt to combat such decontextualisation e.g. by means of a colour plate for fruit given in the back matter or another reference address in the dictionary. In principle the same holds true for what could be termed as *grammatical decontextualisation* in the sense of different derivations of e.g. *-roba* such as *ithoba*, *seroba* and *diroba* that will alphabetically be scattered over the dictionary. For *-roba* this would mean lemmatisation of all derivations in their appropriate alphabetical positions (reference positions) thus decontextualised, to be contextualised by cross-reference to the main lemma *-roba* and its treatment.

The lumping approach surely brings all these derivations together so that they can be treated together and studied as a grammatical set. Contextualisation is further supported by GNSW lemmatising derived forms separately with implicit reference to ROBA. The article of ROBA in GNSW is followed by no less than 51 derivations entered as untreated lemmas cross-referenced to *-roba*:

robagana v. ROBA
robaganedi, se-/di- v. ROBA
robaganelo bo- v. ROBA
(See the appendix for the complete list)

The value of such cross-references for derived forms where all the affixes are suffixes in a stem dictionary is questionable because they all end up alphabetically directly following the article of the lemmatised and treated stem, i.e. ROBA in this case. It does not help the user much if he/she looks up, for example, *robagana* only to be referred to the article of ROBA directly above where (s)he has to work down through the entire user-unfriendly article layout anyway. A more precise reference address within the article of ROBA, i.e. *robagana*

would have been more helpful.

In terms of space utilisation and especially Ziervogel's (1965: 45) claim that the paradigm lumping approach prevents repetition, Van Wyk (1995: 88, 91-2) has shown in a critical review of this dictionary that in following this approach the compilers did *not* manage to avoid repetition. In his view they *introduced* redundancy by having to resort to unnecessary cross-referencing.

This brings no gain in economy compared with word dictionaries. The number of entries is the same for both types, the only difference being the structure and the alphabetic classification of the entries. (Van Wyk 1995: 88)

It also results in overuse of the medio structure.

Should the lexicographer really wish to include entire paradigms of verbal derivations, a splitting approach would be more user friendly: modules 1-32 would be given as main lemmas, each with treatment and will naturally alphabetically be grouped together anyway. Thus there will not be loss in economy and because they will alphabetically be in close proximity, morphological relations would to a large extent be visible and cross-referencing will be limited.

6. Conclusion

The GNSW is the most comprehensive dictionary ever compiled for Sepedi and as such remains an invaluable reference source even after four decades — it is a monument for the language. The GNSW scores high marks as a grammar reference source.

Viewed from many other angles however, GNSW is less effective as a dictionary, especially on different aspects pertaining to lemma selection, user-friendliness and comment on semantics. Initial criticism by sources such as van Wyk (1995) and Prinsloo and De Schryver (1999) were aimed at detrimental aspects of alphabetical ordering and the lemmatisation approach. They concluded among others that stem lemmatisation is the wrong option for a disjunctively written language and that a phonemic ordering is highly problematic from a user perspective.

In this article the selection and presentation of the lemmas were critically evaluated. It is highly unlikely that most of the lemmas will be looked for by target users. The lexicographer should not be creative in the sense of inventing words. He remains a recorder of the language and in the words of Phillip Gove (1961) should not attempt to set its style. (S)he should reflect what is real, the real language as used in print and speech, not that which is possible. Precious dictionary space should rather be used to include more words from the living language than artificially created possible reduplications.

The compilers focused on the completion of grammatical modular paradigms to the extent that the actual existence of most lemmas are questionable as supported by a limited user study, corpus evidence and treatment in other Sepedi dictionaries. Comment on semantics, the most important information

type that users are looking for was grossly neglected. Finally it was argued that the medio structure is largely ineffective.

Note

1. In GNSW *-rōbakantše* and *-rōbakantšwe* incorrectly appear with a circumflex on the *-e*.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85763).

The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.

References

Dictionaries

- GNSW = Ziervogel, D. and P.C. Mokgokong. 1975. *Pukuntšuu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.
- NEND = Kriel, Theunis J. 1976. *The New English–Northern Sotho Dictionary, English–Northern Sotho, Northern Sotho–English*. Johannesburg: Educum Publishers.
- ONSD = De Schryver, G.-M. et al. (Eds.). 2008. *Pukuntšuu ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. Cape Town: Oxford University Press.
- POP = Kriel, T.J., D.J. Prinsloo and B.P. Sathekge. 1997. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Cape Town: Pharos.
- PTLH = Mojela, M.V. (Ed.). 2007. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
- PUKU 1 = Kriel, T.J. 1983. *Pukuntšuu woordeboek*. Pretoria: J.L. van Schaik.
- PUKU 2 = Kriel, T.J. and E.B. van Wyk. 1989. *Pukuntšuu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.
- SLEPD = Mojela, M.V., M.C. Mphahlele, M.P. Mogodi en M.R. Selokela. 2006. *Sesotho sa Leboa/English Pukuntšuu Dictionary*. Cape Town: Phumelela.

Other sources

- Bothma, T.J.D. and D.J. Prinsloo. 2013. Automated Dictionary Consultation for Text Reception: A

- Critical Evaluation of Lexicographic Guidance in Linked Kindle-dictionaries. *Lexicographica* 29(1): 165-198. Berlin: De Gruyter.
- Gouws, Rufus H. and Danie J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS.
- Gove, Philip B. (Ed.)**. 1961. *Webster's Third New International Dictionary of the English Language*. Springfield, Mass.: Merriam-Webster.
- Haas, M.R.** 1962. What Belongs in the Bilingual Dictionary? Householder, F.W. and S. Saporta (Eds.). 1962. *Problems in Lexicography*: 45-50. Bloomington: Indiana University/The Hague: Mouton.
- Hartmann, Reinhard R.K. (Ed.)**. 1983. *Lexicography: Principles and Practice*. London/New York: Academic Press.
- New Words in English*. <http://www.learn-english-today.com/new-words/new-words-in-english.html>. Accessed 12 June 2014.
- Prinsloo, D.J.** 1994. Lemmatization of Verbs in Northern Sotho. *South African Journal of African Languages* 14(2): 93-102.
- Prinsloo, D.J.** 2009. Current Lexicography Practice in Bantu with Specific Reference to the Oxford Northern Sotho Dictionary. *International Journal of Lexicography* 22(2): 151-178.
- Prinsloo, D.J. and G.-M. De Schryver.** 1999. The Lemmatization of Nouns in African Languages with Special Reference to Sepedi and Cilubà. *South African Journal of African Languages* 19(4): 258-275.
- Prinsloo, Danie J., Ulrich Heid, Theo J.D. Bothma and Gertrud Faaß.** 2011. Interactive, Dynamic Electronic Dictionaries for Text Production. Kosem, I. and K. Kosem (Eds.). 2011. *Electronic Lexicography in the 21st Century. New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011*: 215-220. Bled: Trojina.
- PSC = Pretoria Sepedi Corpus. University of Pretoria.
- Van Wyk, E.B.** 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. *Lexikos* 5: 82-96.
- Vintage patterns*. http://vintagepatterns.wikia.com/wiki/Patch_Press_379. Accessed 12 June 2014.
- Wells, Ronald A.** 1973. *Dictionaries and the Authoritarian Tradition. A Study in English Usage and Lexicography*. Janua Linguarum. Series Practica 196. Berlin/The Hague: Mouton de Gruyter.
- Ziervogel, D.** 1965. Probleme van leksikografie in die Suid-Afrikaanse Bantoetale. *Taalfasette* 1: 45-53.
- Ziervogel, D. and P.C. Mokgokong.** 1975. *Pukuntšu ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa-Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho-Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho-Afrikaans/English*. Pretoria: J.L. van Schaik.

Appendix: The verb stem *ROBA* in GNSW

-ritša, bo-	1122	ROBA
<p>waar/waarheen stadig geloop word // place where/to which one walks slowly; moritšetši, ba- (moritšētši) pers. dev.; moritšetšo, me- (moritšētšo) man. dev.; RITŠETŠANA (-ritšētšana, -ritšētšane, -ritšētšanwa, -ritšētšanwe) appl. rec.; baritšetšani (baritšētšani) pers. dev.; moritšetšano, me- (moritšētšano) man. dev.; RITŠISA (-ritšiša, -ritšišitšê, -ritšišwa, -ritšišitšwê) caus.; moritšiši, ba- pers. dev.; moritšišo, me- (moritšišo) man. dev.; RITŠISANA (-ritšišana, -ritšišane, -ritšišanwa, -ritšišanwe) caus. rec.; baritšišani pers. dev.; moritšišano, me- (moritšišano) man. dev.</p>		<p>argument, subdue, be indifferent to what is said, give a hiding; ~ <i>pele</i> ontmoedig, iemand se hart breek // discourage, break one's heart; ~ <i>thokô</i> dom wees // be stupid; <i>thutêlô-bogolo e a</i> ~ dis nie maklik om te leer as jy oud is nie // it is not easy to learn when one has grown old; ~ <i>dikêno</i> belofte verbreek // break one's promise; ~ <i>molaô</i> wet oortree // break/transgress the law; <i>ithôba</i> (-ithôba) refl.; ~ <i>dingalô</i> die vrugte van jou arbeid pluk // reap the fruits of one's labour; morobi, ba- (morôbi) pers. dev.; morobjá, ba- (morôbja) pers. pass. dev.; serobá, di- (serôba) stock, kerie // stick; serobjá, di- (serôbja) impers. pass. dev.; ROBAGANA (-rôbagana, -rôbagane, -rôbaganwa, -rôbaganwe) ass.; in stukke gebreek w. // b. broken into pieces; morobagani, ba- (morôbagani) pers. dev.; morobagano, me- (morôbagano) man. dev.; ROBAGANELA (-rôbagenêla, -rôbaganêtšê, -rôbaganêlwa, -rôbaganêtšwe) appl. < ROBAGANA; morobaganelo (borôbaganêlô) lo. dev.; morobaganelo, me- (morôbaganêlô) man. dev.; serobagane, di- (serôbaganêdi) impers. dev.; ROBAGANTŠHA (-rôbagantšha, -rôbagantšhitšê, -rôbagantšhwa, -rôbagantšhitšwê) caus. < ROBAGANA; in stukke breek // break into pieces; morobagantšhi, ba- (morôbagantšhi) pers. dev.; morobagantšho, me- (morôbagantšhó) man. dev.; ROBAGANTŠHETŠA (-rôbagantšhêtšá, -rôbagantšhêditšê, -rôbagantšhêtšwa, -rôbagantšhêditšwê) caus. appl. < ROBAGANA; morobagantšhetšá, ba- (morôbagantšhêtšá) pers. dev.; morobagantšhetšo, me- (morôbagantšhêtšô) man. dev.; ROBAGANTŠHETŠANA (-rôbagantšhêtšana, -rôbagantšhêtšane, -rôbagantšhêtšanwa, -rôbagantšhêtšanwe) caus. appl. rec. < ROBAGANA; barobagantšhetšani (barôbagantšhêtšani) pers. dev.; morobagantšhetšano, me- (morôbagantšhêtšanô) man. dev.; ROBAGANYA (-rôbaganya, -rôbagantšê, -rôbaganywa, -rôbagantšwê) caus. < ROBAGANA; in stukke breek // break into pieces; morobaganyi, ba- (morôbaganyi) pers. dev.; morobaganyo, me- (morôbaganyô) man. dev.; ROBAGANYETŠA (-rôbaganyêtšá, -rôbaganyêditšê, -rôbaganyêtšwa, -rôbaganyêditšwê) caus. appl. < ROBAGANA; morobaganyetšá, ba- (morôbaganyêtšá) pers. dev.; morobaganyetšo, me- (morôbaganyêtšô) man. dev.; ROBAGANYETŠANA (-rôbaganyêtšana, -rôbaganyêtšane, -rôbaganyêtšanwa, -rôbaganyêtšanwe) caus. appl. rec. < ROBAGANA; barobaganyetšani (barôbaganyêtšani) pers. dev.; morobaganyetšano, me- (morôbaganyêtšanô) man. dev.; ROBAKA (-rôbaka, -rôbakilê, -rôbakwa, -rôbakilwê) it.; <i>ronden-bont</i> breek // break indiscriminately; <i>bošego bo rôbakilwê</i> dit word lig // it is dawning; morobaki, ba- (morôbaki) pers. dev.; morobako, me- (morôbakô) man. dev.;</p>
<p>-ritša, bo- v. RITŠA</p>		
<p>-ritša, se-/di- v. RITŠA</p>		
<p>ritšetša v. RITŠA</p>		
<p>ritšetšana v. RITŠA</p>		
<p>-ritšetšani, ba- v. RITŠA</p>		
<p>-ritšetšano, me-/me- v. RITŠA</p>		
<p>-ritšetši, me-/ba- v. RITŠA</p>		
<p>-ritšetšo, me-/ma- v. RITŠA</p>		
<p>-ritšetšo, me-/me- v. RITŠA</p>		
<p>-ritši, se-/di- v. RITŠA</p>		
<p>ritšiša v. RITŠA</p>		
<p>ritšišana v. RITŠA</p>		
<p>-ritšišani, ba- v. RITŠA</p>		
<p>-ritšišano, me-/me- v. RITŠA</p>		
<p>-ritšiši, me-/ba- v. RITŠA</p>		
<p>-ritšišo, me-/me- v. RITŠA</p>		
<p>-ritšo, me-/me- v. RITŠA</p>		
<p>-ritwá, se-/di- v. RITA</p>		
<p>-RÔ, bo-/ma-, boROFA, klou (hond, leeu) // paw (dog, lion)</p>		
<p>-RÔ, me-/me- (morô) sop, sous // soup, gravy; ~ <i>wa pitsi</i>, <i>se mphišê</i>, <i>ke rilê mola ba go lahla</i>, <i>ka go ja mphagô ka go nwa</i> stank vir dank is die wêreld se loon // get more kicks than halfpence; ~, <i>motsefêla-bašebi</i>, <i>bašomi dipelo tša baba</i> 'n mens moet vir klein dingetjies ook dankbaar wees // one must be thankful for small mercies</p>		
<p>-RÔ, se-/di- ['n soort eetbare bol] // a kind of edible bulb</p>		
<p>-ROBA, le-/ma-, cf. leŠOBA, gat, (pl.) steekwonde // hole, (pl.) stab wounds</p>		
<p>-ROBA, me-/me- (morôba) gemak, ledigheid // comfort, idleness; <i>-bolawa ke</i> ~ op jou dooie gemak werk // work at leisure; <i>-ba morôbêng</i> op jou gemak wees // be at one's leisure</p>		
<p>-ROBA, se-/di- ingang aan agterkant van die lapa // back entrance to lapa</p>		
<p>-robá, se-/di- v. ROBA</p>		
<p>ROBA (-rôba, -rôbilê, -rôbja, -rôbilwê) breek, afbreek, swaar w. // break, break off, b. heavy; ~ <i>monô o tee nege</i> // nine; ~ <i>menô ê mebêdi</i> agt // eight; ~ <i>diêma</i> onkontroleerbaar word, jou ongewoon gedra // become uncontrollable, behave unusually; ~ <i>kgarebê leoto</i> 'n meisie swanger maak // impregnate a girl; ~ <i>molala</i> 'n argument wen, betoel, tem, jou nie steur aan wat gesê word nie, 'n pak slae gee // win an</p>		

ROBA	1123	ROBA
<p>ROBAKANA (-rôbakana, -rôbakane, -rôbakanwa, -rôbakanwe) rond-en-bont gebreek w. // be broken indiscriminately; <i>kgôrông go rôbakanwe</i> die raad het uiteengegaan // the meeting broke up; dirobakani (dirôbakani) impers. dev.; morobakano, me- (morôbakano) man. dev.; ROBAKANTŠHA (-rôbakantšha, -rôbakantšhitšê, -rôbakantšhwa, -rôbakantšhitšwê) caus. < ROBAKANA; rond-en-bont breek // break indiscriminately; morobakantšhi, ba- (morôbakantšhi) pers. dev.; morobakantšho, me- (morôbakantšhó) man. dev.; serobakantšhwá, di- (serôbakantšhwa) impers. pass. dev.; ROBAKANTŠHETSÁ (-rôbakantšhêtša, -rôbakantšhêditšê, -rôbakantšhêtšwa, -rôbakantšhêditšwê) caus. appl. < ROBAKANA; morobakantšhetši, ba- (morôbakantšhetši) pers. dev.; morobakantšhetšo, me- (morôbakantšhêtšó) man. dev.; ROBAKANTŠHETSÁNA (-rôbakantšhêtšana, -rôbakantšhêtšane, -rôbakantšhêtšanwa, -rôbakantšhêtšanwe) caus. appl. rec. < ROBAKANA; barobakantšhetšani (barôkantšhêtšani) pers. dev.; morobakantšhetšano, me- (morôbakantšhêtšanó) man. dev.; ROBAKANYA (-rôbakanya, -rôbakantšê, -robakanywa, -robakantšwê) caus. < ROBAKANA; rond-en-bont breek // break indiscriminately; morobakanyi, ba- (morôbakanyi) pers. dev.; morobakanyo, me- (morôbakanyó) man. dev.; serobakanywá, di- (serôbakanywa) impers. pass. dev.; ROBAKANYETSÁ (-rôbakanyêtša, -rôbakanyêditšê, -rôbakanyêtšwa, -rôbakanyêditšwê) caus. appl. < ROBAKANA; morobakanyetši, ba- (morôbakanyetši) pers. dev.; morobakanyetšo, me- (morôbakanyêtšó) man. dev.; ROBAKANYETSÁNA (-rôbakanyêtšana, -rôbakanyêtšane, -rôbakanyêtšanwa, -rôbakanyêtšanwe) caus. appl. rec. < ROBAKANA; barobakanyetšani (barôbakanyêtšani) pers. dev.; morobakanyetšano, me- (morôbakanyêtšanó) man. dev.; ROBEGA (-robêga, -robêgilê) neutr.; breekbaar w. // b. breakable, b. fragile; -<i>otlolla mohlašana e sa le ô monana, wa gola o tlô</i> ~ jy moet die boompie buig solank hy nog jonk is // as the twig is bent so the tree grows; <i>pelo e robêgilê</i> gekwets raak, seerkry (fig.) // get hurt (fig.); morobegi, ba- (morôbêgi) pers. dev.; 'n persoon van wie 'n liggaamsdeel gebreek is // a person whose limb is broken; morobego, me- (morôbêgô) man. dev.; thobego, (n-)/di- (thôbêgo) man. dev.; breekbaarheid // breakability; †beenbreuk // †fracture; ROBEGETSÁ (-rôbêgêtša, -rôbêgêditšê, -rôbêgêtšwa, -rôbêgêditšwê) v. ROBESETSÁ; ~ <i>lehlôkwana tsêbêng</i> raad gee // advise; ~ <i>lerapô têng</i> iemand swaar straf // punish someone severely; morobegetši, ba- (morôbêgêtši) pers. dev.; morobegetšo, me- (morôbêgêtšó) man. dev.; ROBELA (-rôbêla, -rôbêtše, -rôbêlwa, -rôbêtšwe) appl.;</p>	<p>~ <i>lerapô têng</i> iemand hartseer maak // break someone's heart; ithôbela (-ithôbêla) refl.; ~ <i>legong hlôgông</i> jouself in die moelikheid laat beland // bring trouble upon oneself; morobedi, ba- (morôbêdi) pers. dev.; morobelo, me- (morôbêlô) man. dev.; ROBELANA (-rôbêlana, -rôbêlane, -rôbêlanwa, -rôbêlanwe) appl. rec.; barobelani (barôbêlani) pers. dev.; morobelano, me- (morôbêlanó) man. dev.; ROBESETSÁ (-rôbêsêtša, -rôbêsêditšê, -rôbêsêtšwa, -rôbêsêditšwê), cf. ROBEGETSÁ, binne-in afbreek, afbreek sodat 'n stuk agterbly // break off inside, break off in such a way that a piece remains behind; morobesetši, ba- (morôbêsêtši) pers. dev.; morobesetšo, me- (morôbêsêtšó) man. dev.; ROBIŠA (-rôbiša, -rôbišitšê, -rôbišwa, -rôbišitšwê) caus.; laat/help breek, ens. // help/cause to break, etc.; morobiši, ba- (morôbiši) pers. dev.; morobišo, me- (morôbišó) man. dev.; ROBIŠANA (-rôbišana, -rôbišane, -rôbišanwa, -rôbišanwe) caus. rec.; barobišani (barôbišani) pers. dev.; morobišano, me- (morôbišanó) man. dev.; ROBOKA (-rôbôka, -rôbôkilê, -rôbôkwa, -rôbôkilwê) it., v. ROBAKA; moroboki, ba- (morôbôki) pers. dev.; moroboko, me- (morôbôkô) man. dev.; ROBOKANA (-rôbôkana, -rôbôkane, -rôbôkanwa, -rôbôkanwe) rec. < ROBOKA; v. ROBAKANA; dirobokani (dirôbôkani) pers. dev.; morobokano, me- (morôbôkanó) man. dev.; ROBOKANTŠHA (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < ROBOKANA; (-rôbôkantšha, -rôbôkantšhitšê, -rôbôkantšhwa, -rôbôkantšhitšwê) caus. < ROBOKANA; v. ROBAKANTŠHA; morobokantšhi, ba- (morôbôkantšhi) pers. dev.; morobokantšho, me- (morôbôkantšhó) man. dev.; serobokantšhwá, di- (serôbôkantšhwa) impers. pass. dev.; ROBOKANTŠHETSÁ (-rôbôkantšhêtša, -rôbôkantšhêditšê, -rôbôkantšhêtšwa, -rôbôkantšhêditšwê) caus. appl. < ROBOKANA; morobokantšhetši, ba- (morôbôkantšhêtšhi) pers. dev.; morobokantšhetšo, me- (morôbôkantšhêtšó) man. dev.; ROBOKANTŠHETSÁNA (-rôbôkantšhêtšana, -rôbôkantšhêtšane, -rôbôkantšhêtšanwa, -rôbôkantšhêtšanwe) appl. rec. < ROBOKANTŠHA; barobokantšhetšani (barôbokantšhêtšani) pers. dev.; morobokantšhetšano, me- (morôbôkantšhêtšanó) man. dev.; ROBOKANYA (-rôbôkanya, -rôbôkantšê, -rôbôkanywa, -rôbôkantšwê) caus. < ROBOKANA; v. ROBAKANYA; morobokanyi, ba- (morôbôkanyi) pers. dev.; morobokanyo, me- (morôbôkanyó) man. dev.; serobokanywá, di- (serôbôkanywa) impers. pass. dev.; ROBOKANYETSÁ (-rôbôkanyêtša, -rôbôkanyêditšê, -rôbôkanyêtšwa, -rôbôkanyêditšwê) appl. < ROBAKANYA; morobokanyetši, ba- (morôbôkanyetši) pers. dev.; morobokanyetšo, me- (morôbôkanyêtšó) man. dev.</p>	

-robádi, bo-

1124

ROBALA

dev.; *ROBOKANYETŠANA* (-róbókanyētšana, -róbókanyētšane, -róbókanyētšani, -róbókanyētšani) appl. rec. < *ROBOKANYA*; barobokanyetšani (baróbókanyētšani) pers. dev.; morobokanyetšano, me- (moróbókanyētšano) man. dev.

-robádi, bo- v. ROBALA

-robadi, mo-/ba- v. ROBALA

-robadi, se-/di- v. ROBALA

-ROBADILÉPE, mo-/me- (moróbadilepê) doppruimboom // Pappia fulva

robadiša v. ROBALA

robadišana v. ROBALA

-robadišani, ba- v. ROBALA

-robadišano, mo-/me- v. ROBALA

-robadiši, mo-/ba- v. ROBALA

-robadišo, mo-/me- v. ROBALA

-robadišwá, mo-/ba- v. ROBALA

robagana v. ROBA

-robagane, se-/di- v. ROBA

robaganela v. ROBA

-robaganelo, bo- v. ROBA

-robaganelo, mo-/me- v. ROBA

-robagani, mo-/ba- v. ROBA

-robagano, mo-/me- v. ROBA

robagantšha v. ROBA

robagantšhetša v. ROBA

robagantšhetšana v. ROBA

-robagantšhetšani, ba- v. ROBA

-robagantšhetšano, mo-/me- v. ROBA

-robagantšhetši, mo-/ba- v. ROBA

-robagantšhetšo, mo-/me- v. ROBA

-robagantšhi, mo-/ba- v. ROBA

-robagantšho, mo-/me- v. ROBA

robaganya v. ROBA

robaganyetša v. ROBA

robaganyetšana v. ROBA

-robaganyetšani, ba- v. ROBA

-robaganyetšano, mo-/me- v. ROBA

-robaganyetši, mo-/ba- v. ROBA

-robaganyetšo, mo-/me- v. ROBA

-robaganyi, mo-/ba- v. ROBA

-robaganyo, mo-/me- v. ROBA

robaka v. ROBA

robakana v. ROBA

-robakani, ba- v. ROBA

-robakano, mo-/me- v. ROBA

robakantšha v. ROBA

robakantšhetša v. ROBA

robakantšhetšana v. ROBA

-robakantšhetšani, ba- v. ROBA

-robakantšhetšano, mo-/me- v. ROBA

-robakantšhetši, mo-/ba- v. ROBA

-robakantšhetšo, mo-/me- v. ROBA

-robakantšhi, mo-/ba- v. ROBA

-robakantšho, mo-/me- v. ROBA

-robakantšhwá, se-/di- v. ROBA

robakanya v. ROBA

robakanyetša v. ROBA

robakanyetšana v. ROBA

-robakanyetšani, ba- v. ROBA

-robakanyetšano, mo-/me- v. ROBA

-robakanyetši, mo-/ba- v. ROBA

-robakanyetšo, mo-/me- v. ROBA

-robakanyi, mo-/ba- v. ROBA

-robakanyo, mo-/me- v. ROBA

-robakanywá, se-/di- v. ROBA

-robaki, mo-/ba- v. ROBA

-robako, mo-/me- v. ROBA

ROBALA (-róbala, -róbétše, -róbala, -róbétšwe) slaap, lê, geslagtelik verkeer // sleep, lie down, have sexual intercourse; *o robétše* hy is aan die slaap // he is asleep; *go mo róbala* geslagtelik met haar verkeer // have sexual intercourse with her; ~ *borókô* nie uitgeslape genoeg wees nie, gemaklik lewe // not to be shrewd enough, live comfortably; ~ *botala* kaal slaap, geen komberse om onder te slaap hê nie // sleep naked, have no blankets to sleep under; ~ *ka leihlô le tee* voorgee dat jy slaap, voorgee dat jy dom is // pretend to go and sleep, pretend to be stupid; ~ *le dikgogo* saam met die hoenders gaan slaap // go to bed early; ~ *le dipudi* baie onnet wees // be very untidy; ~ *le letšatši* v. ~ *le dikgogo*; ~ *le matôlô* sonder 'n vroumens slaap // sleep without a woman; ~ *le taba* 'n geheim bewaar, eerste kennis dra van // keep a secret, first to know; ~ *phokéng* swaarkry // suffer; ~ *sepatikane* 'n dutjie neem voordat jy gaan slaap // take a nap before going to bed; ~ *sebutla/semmutla* met oop oë slaap // sleep with eyes open; ~ *seróbalela* v. *serobalela*; ~ *sethaka* saam op een plek slaap soos seuns van dieselfde ouderdom // sleep together in one place as boys of the same age-group do; *naga e sa robétše* die mense is nog onderontwikkeld, vredenstyd // people are still under-developed, peace time; *ke ilô* ~ *le wêna kobô e tee* ek sal jou agtervolg // I shall follow you; *borobadi* (boróbadi) slaperigheid, slaapsiekte // sleepiness, sleeping sickness; *borobalo*, ma- (boróbalô) lo. dev.; slaapplek, akkommodasie // sleeping place, accommodation; -*kgopéla maróbalô* slaapplek vra // apply for accommodation; *morobadi*, ba- (moróbadi) pers. dev.; slaper // sleeper; *morobalo*, me- (moróbalô) man. dev.; *morobalwá*, ba- (moróbalwa) pers. pass. dev.; 'n vrou met wie geslagtelik verkeer w. // a woman with whom one has sexual intercourse; *serobadi*, di- (seróbadi) impers. dev.; *serobalo*, di- (seróbalô) kombers // blanket; *thobalo*, (n-)/di- (thóbalô) man. dev.; slaap, geslaap // sleep, sleeping; *ROBADIŠA* (-róbadiša, -róbadišitšê, -róbadišwá, -róbadišitšwê) caus.; cf. *ROBATŠA*, aan die slaap maak // cause to sleep; ~ *kgogo borókô* kul // deceive; ~ *mmutla* kul // deceive; ~ *nôga ya mpa* 'n bietjie eet // eat a little; *morobadiši*, ba- (moróbadiši) pers. dev.; *morobadišo*, me- (moróbadišô) man. dev.; *morobadišwá*, ba- (moróbadišwa) pers. pass. dev.; persoon wat aan die slaap gemaak word // a person who is caused/helped to sleep; *thobadišo*, (n-)/di- (thó-badišô) man. dev.; *ROBADIŠANA* (-rób-

Lexicographic Treatment of Kinship Terms in an English/Sepedi–Setswana–Sesotho Dictionary with an Amalgamated Lemmalist

D.J. Prinsloo, *Department of African Languages, University of Pretoria, Pretoria, South Africa (danie.prinsloo@up.ac.za)*

Abstract: This article describes the lemmatisation and treatment of kinship terms in a proposed English–Sotho, Sotho–English dictionary with an amalgamated lemmalist. The first requirement is to build a list of kinship terminology for the Sotho languages. Secondly, it is necessary in terms of space restriction to determine the most frequently used forms to be lemmatised in such a dictionary. Thirdly, the macrostructure and microstructure of the dictionary should be planned in terms of an amalgamated approach. A short explanation of the amalgamated model will be presented and a schematic illustration of the paternal family tree structure in the Sotho languages is given in the appendix. Specific attention is given to the compilation of the amalgamated lemmalist focusing on absolute cognates and absolute cognates with a difference in form. Finally, where the reduction of huge quantities of terms, e.g. all derived forms of a specific term in all three Sotho languages are at stake, a lexicographic convention will be suggested to sensibly reduce the number of lemmas and to combat redundancy.

Keywords: AMALGAMATED DICTIONARIES, KINSHIP TERMS, SOTHO LANGUAGES, LEXICOGRAPHIC CONVENTION, CLOSELY RELATED LANGUAGES

Opsomming: Die leksikografiese hantering van verwantskapsterme in 'n Engels/Sepedi–Setswana–Sesotho-woordeboek met 'n gealgameerde lemmalys. Hierdie artikel beskryf die lemmatisering en bewerking van verwantskapsterme in 'n voorgestelde Engels–Sotho, Sotho–Engels woordeboek met 'n gealgameerde lemmalys. Die eerste vereiste is die samestelling van 'n lys van verwantskapsterminologie vir die Sothotale. Tweedens is dit nodig om ter wille van ruimtebesparing die mees gebruiklike vorme te bepaal wat in so 'n woordeboek gelemmatiseer moet word. Derdens moet die makro- en mikrostruktuur van die woordeboek beplan word in terme van 'n gealgameerde benadering. 'n Kort verduideliking van die gealgameerde model sal aangebied word en 'n skematiese voorstelling van die paterne stamboomstruktuur in die Sothotale word in die bylaag aangegee. Spesifieke aandag word gegee aan die samestelling van die gealgameerde lemmalys met die fokus op absolute kognate en absolute kognate met 'n vormverskil. Ten slotte, waar die vermindering van groot hoeveelhede van die terme, byvoorbeeld alle afgeleide vorme van 'n spesifieke term in al drie Sothotale ter sake is,

sal 'n leksikografiese konvensie voorgestel word om die aantal lemmas sinvol te verminder en redundansie te bestry.

Slutelwoorde: GEAMALGAMEERDE WOORDEBOEKE, VERWANTSKAPSTERME, SOTHOTALE, LEKSIKOGRAFIESE KONVENSIE, NOUVERWANTE TALE

1. Introduction

The aim of this article is to describe the treatment of kinship terms in an English–Sotho, Sotho–English dictionary with an amalgamated lemmalist. The kinship system in the Sotho languages is complicated, (see appendix), and was selected as an object of study in order to test the viability of the amalgamated approach for such complex structures. Prinsloo (2012) distinguishes three categories of kinship terms for Sepedi, i.e. underived single words such as *malome* 'uncle', *rakgadi* 'aunt' and *tate* 'father', derived words such as *malomeagwe* 'his uncle', *morwediate* 'my daughter' and *bomalomeago* 'your uncles' and phrases such as possessive constructions *mogatša wa mokgotse wa ka* 'my brother in law's wife'. He suggested specific lemmatisation strategies to cater for the large number of kinship terms in these categories in Sepedi, including a specific dictionary convention. An attempt to handle kinship terminology for three languages simultaneously is an even greater challenge since quantity wise the number of kinship terms to be lemmatised is threefold and new challenges on macrostructural as well as on microstructural levels come to the fore. The question is whether it is possible to do justice to all three languages in terms of similarities versus differences, following an amalgamated approach.

The first requirement for the lexicographer is to build a list of kinship terminology for the Sotho languages. It is also necessary in terms of space restriction to determine the most frequently used forms to be lemmatised in such a dictionary. In this article an attempt will be made to collect a number of kinship terms for the Sotho languages, i.e. Sepedi, Setswana and Sesotho. Secondly, the frequency of use of Sotho kinship terms in corpora for these languages will be determined. Thirdly, the treatment of Sotho kinship terms in separate English Sepedi/Setswana/Sesotho dictionaries will be studied in order to establish the viability of such an amalgamated approach and in order to suggest model dictionary articles.

The collection of Sepedi kinship terminology is mainly based on Prinsloo and Van Wyk (1992), Setswana on Van Wyk and Haasbroek (1990) and Sesotho on Molalapata (2004) supplemented by terms found in dictionaries and corpora of the Sotho languages. By way of introduction, a short explanation of the amalgamated model will be presented, followed by a schematic illustration of the paternal family tree structure in the Sotho languages. Finally, the formulation of model entries with an amalgamated approach will be presented.

2. The amalgamated model

The design of amalgamated dictionaries is credited to Martin and Gouws (2000) for introducing the concept and also for compiling the first amalgamated dictionary for Afrikaans and Dutch, *Groot Woordeboek Afrikaans en Nederlands* (ANNA).

The ANNA-approach is to provide treatment for what the amalgamated languages have in common first (A|N, A=Afrikaans, N=Nederlands (Dutch)) followed by the treatment of aspects applicable to the specific languages. Consider the article of *ouderwets* 'old fashioned' from ANNA. The article consists of three sections viz. A|N, N and A. Similarities and differences are indicated throughout by the symbols "=" 'equal' and "≠" 'differ' respectively.

ouderwets b.nw., *ouderwets* b.nw.

A|N (**v. vroeger**) *ouderwets* = ouderwetse kleren *ouderwetse klere*; een ouderwetse stoomtrein 'n *ouderwetse stoomtrein*; ouderwetse opvattingen *ouderwetse opvattinge*; hopeloos ouderwets *hopeloos ouderwets* ≠ *stewige ouderwetse meubels* oerdegelyk meubilair

N (**net als vroeger**) *outyds*, *ouwêrelds* = ouderwetse degelykheid *outydse deeglykheid*; een ouderwetse winter 'n *outydse winter* ≠ het was weer ouderwets gezellig *dit was weer gesellig soos in die ou tyd*

A (**oulik; slim**) *bijdehand* = 'n *ouderwetse kind* een bijdehand kind

Detailed discussions of the amalgamated approach and of ANNA in particular can be found in Martin (2012a and 2012b), Martin and Gouws (2000), Marais (2011), Bosman (2013) and in the user's guide of ANNA. Martin's intention with the amalgamated model was also to pave the way for other closely related languages:

the aim was not only to produce a contrastive dictionary Afrikaans-Dutch, but also to lay the foundation for an *exportable model*, one that could be used for other closely related languages, such as the 'black' languages in South Africa: Xhosa and Zulu, and North-Sotho, South Sotho and Tswana etc. (Martin 2012b: 413)

Amalgamated dictionaries would employ a single lemmalist for closely related languages such as Afrikaans/Dutch, Sepedi/Setswana/Sesotho, isiZulu/isiXhosa/Siswati/isiNdebele and have a unique microstructural architecture in their treatment of the languages in question.

The first requirement for an amalgamated approach is that the languages to be treated should be closely related, i.e. that they should have a substantial number of words in common.

it can only be applied to closely related languages ... both the 'form' of the words (spelling) needs to be the 'same' and at least one of the meanings. ... there has to be a sufficient critical mass. (Martin 2012b: 414)

Martin (2012b: 415) puts the overlap between Afrikaans and Dutch as 2/3, i.e. 66.7%.

For the Sotho languages Prinsloo compared the 10,000 most frequently used words in Sesotho, Setswana and Sepedi corpora and came to the conclusion that the vocabulary of these languages overlap to a large extent. The three languages have 19,4% words in common, Sepedi and Setswana share 32,7%, Sepedi and Sesotho 26,9% and Setswana and Sesotho 34,4%.

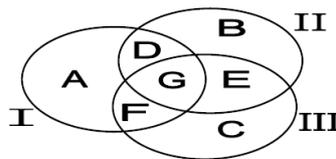


Figure 1: Comparison of the Sotho languages

I = Sepedi; II = Setswana; III = Sesotho

A = 5 978; B = 5 226; C = 5 813; D = 1 333; E = 1 498; F = 746; G = 1 943

(Prinsloo 2005)

This degree of overlap would result in a single amalgamated lemmalist of 22,537 in contrast to a list of 30,000 lemmas (10,000 each for Sepedi, Setswana and Sesotho) if three separate dictionaries were compiled, thus a saving of almost 30%. For the English–Sesotho section the space saving stands at 67% compared to three English sections in three separate dictionaries, i.e. English–Sepedi, English–Setswana and English–Sesotho.

Martin (ANNA: 25) distinguishes five different types of words relevant to an amalgamated approach, i.e. (a) absolute cognates: words in the related languages which are identical in form and meaning, (b) absolute cognates with difference in form, (c) partial cognates: words that differ in at least one sense, (d) non-cognates: words with the same meaning but clear difference in form and (e) false friends: words identical in form but which differ in meaning.

In this article the focus will be on absolute cognates, and cognates with a difference in form.

Absolute cognates are most beneficial to an amalgamated approach because a single lemma represents all of the related languages. For example, the translation equivalents for *woman*, *love* and *neck* in Sepedi, Setswana and Sesotho are identical.

woman *n.* mosadi

love *v.* rata

neck *n.* molala

Likewise, *efe* in all three Sotho languages can be translated with a single equivalent:

efe *enum.* which (one)?

Absolute cognates are discussed in more detail in paragraph 4 below.

Absolute cognates with difference in form, e.g. *kgaetšedi* (Sepedi), *kgaitsadi* (Setswana) and *kgaitsedi* (Sesotho), 'sister/brother' also fit within an amalgamated approach but with some consequences for user-friendliness, cross-referencing and redundancy which will be discussed in more detail below. Partial cognates find their place in an amalgamated approach but require separate treatment for senses where they differ. Non-cognates do not bring much gain in an amalgamated approach since they have to be lemmatised and treated separately. See Prinsloo (2013) for a detailed discussion.

3. Kinship terms in the Sotho languages

An attempt was made to capture single-word kinship terms for Sepedi, Setswana and Sesotho from Prinsloo and Van Wyk (1992), Van Wyk and Haasbroek (1990) and Molalapata (2004) respectively. Their occurrence in the respective corpora was subsequently determined. Finally, a randomly selected number of dictionaries for each of these languages were studied in terms of their lemmatisation and treatment of kinship terms occurring more than once in these corpora.

3.1 Sepedi kinship terms

Single-word kinship terms from Prinsloo and Van Wyk (1992) that occur in the Pretoria Sepedi Corpus (PSC) and in one or more of five randomly selected Sepedi dictionaries are given in table 1 with their frequency counts and inclusion versus omission from the dictionaries marked as "√" and "x" respectively.

Table 1: Sepedi kinship terms

Groot Noord-Sotho-woordeboek, (GNSW), *Sesotho sa Leboa/English Pukuntšu Dictionary* (SLEPD), *Oxford Bilingual School Dictionary: Northern Sotho and English* (ONSD), *Pukuntšutlhaloši ya Sesotho sa Leboa* (PTLH), *Pukuntšu Dictionary* (PUKU1) and *Popular Northern Sotho Dictionary* (POP).

Term	Translation	PSC	GNSW	SLEPD	ONSD	PTLH	PUKU 1	POP
morwa	son	3803	√	√	√	√	√	√
tate	father	1521	√	X	√	√	√	√
mma	mother	1060	√	X	√	√	√	√
morwedi	daughter	1000	√	√	√	√	√	√
ngwetši	daughter in law	637*	√	√	√	√	√	√
kgaetšedi	brother/sister	635	√	√	√	√	√	√
malome	uncle	558	√	√	√	√	√	√
rangwane	father's younger brother	354	√	√	√	√	√	√
rakgadi	father's sister	348	√	√	√	√	√	√
mokgonyana	son in law	260	√	√	√	√	√	√

koko	grandmother	223	√	√	√	√	√	√
rakgolo	grandfather	190	√	√	√	√	√	√
mokgotse	brother/sister in law	189	√	√	√	√	√	√
mogatšaka	my wife/hus- band	188	X	X	√	X	X	√
motlogolo	grandchild, my sister's child	177	√	√	√	√	√	√
tata	father	169	√	√	X	X	√	√
morwarre	brother	127	√	√	√	√	√	√
motswala	cousin	120	√	√	√	√	√	√
moratho	younger brother or a sister	114	√	√	√	√	√	√
ramogolo	father's older brother	114	√	√	√	√	√	√
mme	mother	5524*	√	√	√	X	√	√
molamo	brother in law	315*	√	√	√	√	√	√
mogolle	older brother/ sister	90	√	X	√	X	√	√
mogadibo	brother's wife	82	√	√	√	√	√	√
mmatswale	mother in law	75	√	√	√	√	X	√
mmane	mother's younger sister	61	√	√	√	√	√	√
mogaditšong	co-wife	61	√	√	X	√	√	√
mmamogolo	mother's older sister	57	√	√	X	√	√	√
rra	father	55	X	√	X	X	√	√
morwake	my son	45	X	√	X	√	X	√
samma	younger brother or sister	40	√	√	X	√	√	√
mmangwane	aunt	39	√	√	X	√	√	√
kgaitšedi	brother/sister	38	√	X	X	X	√	√
ngwanangwanaka	my grandchild	35	X	√	X	√	X	X
morwediake	my daughter	34	X	√	X	√	X	X
monyana	younger brother or sister	28	√	X	X	√	√	√
ratswale	father in law	26	√	√	X	√	√	√
matswale	mother in law	14	√	X	X	X	X	√
motswalake	my cousin	14	X	X	X	X	X	√
mogadikane	co-wife	12	√	√	X	√	√	√
mmamalome	uncle's mother	11	X	X	X	X	X	X
tatemogolo	grandfather	9	X	X	X	X	X	X
nnake	my younger brother/sister	8	√	√	X	√	√	X
rakgolokhukhu	great grand- father	8	√	√	X	√	√	X
morwaka	my son	7	X	X	X	X	X	√
morwediaka	my daughter	7	X	X	X	X	X	X
ngwanangwanake	my grandchild	5	X	X	X	X	X	X
mmakgolo	grandmother	3	X	X	X	X	√	X
mogadikana	co-wife	3	X	√	X	√	X	X

* frequency counts include homonyms which are not kinship terms

3.2 Setswana kinship terms

Single-word kinship terms from Van Wyk and Haasbroek (1990) that occur in the

Pretoria Setswana Corpus (PSETC) and in one or more of five randomly selected Setswana dictionaries are given in table 2.

Table 2: Setswana kinship terms

SESD = *Setswana English Setswana Dictionary* TYMYS = *Tlhalosi ya medi ya Setswana*
 TYS = *Thanodi ya Setswana: Sefala kgobokgobo* DSEA = *Dikišinare ya Setswana–English–Afrikaans*
 CSD = *Compact Setswana Dictionary* *Dictionary/Woordeboek*

		PSETC	SESD	TYMYS	TYS	DSEA	CSD
mme	mother	33211*	√	√	√	√	√
nna	I	24685	√	√	X	√	√
rre	father	3677	√	√	√	√	X
mma	mother	3175	√	√	√	√	x
rra	father	2760	√	√	√	√	√
morwa	son	2210	√	√	√	√	√
ngwanaka	my child	1128	√	X	√	√	X
koko	grandmother	714	X	√	√	X	X
malome	uncle	551	√	√	√	√	√
ntate	father	365	√	√	√	√	√
ngwanake	my child	351	√	X	√	√	X
mohumagadi	chief's wife	326*	√	√	√	X	X
mogolole	my elder sibling	272	√	√	√	√	X
nkoko	grandmother	262	√	√	√	√	X
mogatsa	spouse	253	√	√	√	√	√
ngwetsi	daughter-in-law	229	√	√	√	√	X
rremogolo	grandfather	220	√	√	√	√	√
morwadi	daughter	219	√	√	√	X	√
nkgonne	elder brother/sister	217	√	√	√	√	X
rangwane	uncle	217	√	√	√	√	√
rakgadi	aunt	176	√	√	√	√	√
mogatsaka	my spouse	169	√	√	√	√	X
mogatsake	my spouse	169	√	√	√	√	X
ntsala	cousin	168	√	√	√	X	X
ausi	older sister	153	√	√	√	X	X
mmangwane	aunt	153	√	√	√	√	√
morwaaka	my son	139	√	√	√	√	X
morwaake	my son	118	√	√	√	√	X
motlogolo	nephew	109	√	√	√	√	X
setlogolo	grandchild	103	√	√	√	x	√
kgaitsadi	brother/sister	90	√	√	√	√	√

nnake	my younger brother	90	√	√	√	√	√
mmemogolo	grandmother	72	√	√	√	√	X
kgaitsadiake	my brother/sister	71	X	√	√	√	X
mogadibo	my husband's sister/my brother's wife	68	X	√	√	√	X
mogwe	son-in-law, brother-in-law	66	√	√	√	X	√
ntatemogolo	grandfather	62	X	√	√	√	√
morwadiake	my daughter	43	√	√	√	√	X
sesi	sister	39	x	√	X	X	√
aubuti	older brother	38	√	√	√	X	√
morwadiaka	my daughter	36	X	√	√	√	X
ntsalake	my cousin	35	√	√	√	√	X
khumagadi	wife of the king	32	√	√	√	√	X
matsale	mother-in-law	22	√	√	√	X	√
leitibolo	firstborn	21	X	√	√	X	√
mmane	mother's younger sister	20	√	√	√	X	X
mmamalome	uncle's mother	14	X	√	√	X	X
mokgonyana	son in law	13	X	X	√	X	X
motswala	cousin	12	X	X	X	X	X
mogadikane	co-wife	10	√	√	X	X	X
mokgwenyana	son in law	8	X	√	X	X	X
mmamogolo	aunt	7	√	√	X	X	√
mogwagadi	man's father/mother-in-law	7	√	√	X	X	√
ntataago	your father	7	X	√	X	√	X
ratswale	father	7	X	X	X	X	√
rramogolo	grandfather	7	√	X	X	X	√
kgantsadi	brother/sister	4	√	√	X	X	X
mogwagwadi	man's father/mother-in-law	4	X	√	X	X	X
sebare	brother-in-law	4	x	√	X	X	√
ratsale	father	2	x	x	X	X	√

* frequency counts include homonyms which are not kinship terms

3.3 Sesotho kinship terms

Single-word kinship terms from Molalapata (2004) that occur in the Pretoria Sesotho Corpus (PSSC) and in one or more of five randomly selected Sesotho dictionaries are given in table 3.

Table 3: Sesotho kinship terms

NSSD = New South Sotho English Dictionary SSED = Southern Sotho English Dictionary
 LSS = Longman Sethantšo sa Sesotho BUKAN = Bukantswe <http://bukantswe.sesotho.org/>
 FREEL = <http://www.freelang.net/dictionary/sesotho.php>

		PSSC	NSSD	SSED	LSS	BUKAN	FREEL
mme	mother/(and)	19560*	√	X	X	X	√
mora	son	3463	√	√	X	X	√
monna	husband/(man)	3036	√	√	X	√	√
ntate	father	2007	√	√	√	√	√
moradi	daughter	732	√	√	X	√	√
kgaitsedi	brother/sister	244	√	√	X	√	√
ngwetsi	sister-in-law/daughter-in-law	201	√	X	X	√	√
rangwane	uncle	192	X	X	X	√	√
malome	uncle	188	√	√	√	√	√
mmangwane	aunt	141	√	X	X	√	√
moholwane	older sister/brother	133	X	X	X	√	√
nnake	my little sister	90	X	√	X	X	X
rakgadi	aunt	88	X	X	X	√	√
motswala	cousin	67	X	√	X	√	√
moena	younger brother/sister	48	X	√	X	√	√
setlohlo	grandchild	47	√	X	X	√	√
kgorula	youngest child	46	X	X	X	√	√
motjhana	nephew/niece	43	X	√	X	√	√
ntatemoholo	grandfather	43	√	√	√	√	√
mohwehadi	mother-in-law	33	X	√	X	√	√
mohwe	father-in-law	27	√	√	X	√	√
rangoane	uncle	15	X	X	√	X	X
matsale	mother-in-law	14	√	√	√	√	√
molamo	brother	4	X	X	√	X	X

* frequency counts include homonyms which are not kinship terms

From tables 1 to 3, although they do not reflect the full scope of kinship terms and their derivations, it is clear that the lexicographer has to deal with quite a number of kinship terms in each of the Sotho languages. The question is whether an amalgamated approach will be able to do justice at least to the frequently used terms on both macrostructural and microstructural levels.

4. Macrostructural considerations

On the English side of the proposed bi-directional dictionary, a single lemma-

list is given instead of three English lemmalists for three separate English–Sepedi, English–Setswana and English–Sesotho dictionaries. Once again a 67% reduction is possible because the English lemmalist of kinship terms is presented only once. The challenge, however, is the compilation of the lemmalist on the Sotho side where the model requires amalgamation of the three separate lemmalists for Sepedi, Setswana and Sesotho into a single lemmalist. The five types of cognates identified in ANNA have different implications for the model and the first two are considered here. As briefly stated above, *absolute cognates* are the most beneficial because a single lemma can represent all three languages. Consider also the following frequently used and identical absolute cognates in the three Sotho languages.

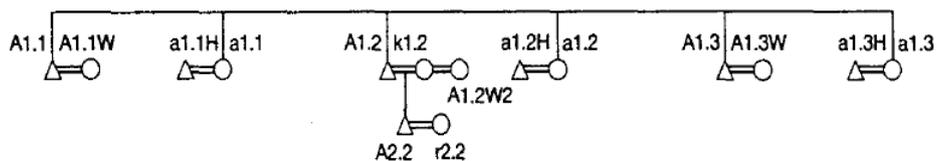
- malome** mother's younger brother
- ntate** (my) father
- rangwane** father's younger brother

In all three Sotho languages *ntate*, *rangwane* and *malome*, respectively, have the same meanings, e.g. *ntate* 'father' in all three languages. Thus in the Sotho to English section space is saved in comparison to three separate dictionaries. *Ramogolo*, *rangwane* and *malome* are lemmatised once instead of three times, each in a Sepedi–English, Setswana–English and Sesotho–English dictionary.

In the English–Sotho section *uncle* is lemmatised only once instead of three times in three separate dictionaries (English–Sepedi, English–Setswana and English–Sesotho).

- uncle** ramogolo/rremogolo (father's older brother), rangwane (father's younger brother), malome (mother's brother)

Ideally a single term for *uncle* in all three of the Sotho languages would have resulted in additional space saving as in the case of *ntate* '(my) father'. In the case of *uncle* semantic divergence does not lie on the level of differences between the Sotho languages — *ramogolo* 'father's older brother' *rangwane* 'father's younger brother' and *malome* 'mother's brother' have the same meanings respectively in all three Sotho languages. They, however, refer to different relations in terms of the age of the related person and his position in the family tree, consider extracts from the family tree from Prinsloo and Van Wyk (1992) given in the appendix:



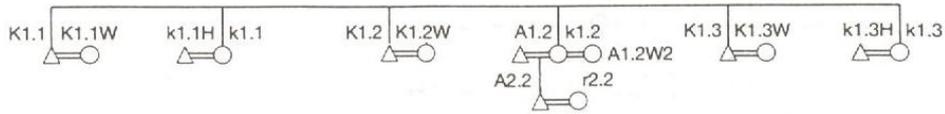


Figure 2: Relationship between a man (A2.2) and his uncles and aunts

Uncle, referring to a man's father's elder brother (A1.1) is *ramogolo*, father's younger brother (A1.3) is *rangwane*. *Uncle*, in reference to mother's younger brother (K1.3) is *malome*. *Ramogolo*, *rangwane* and *malome* are lexicalised terms that could, for lack of equivalents, at best in English be described by means of a paraphrase "a man's father's elder brother", "father's younger brother" and "mother's younger brother".

In the case of *absolute cognates with difference in form*, the first consideration is the presumed knowledge of the target users. The more knowledgeable they are of one or more of the Sotho languages, (a) the more user-friendly an amalgamated lemmalist will be to them, (b) the less problematic it will be for the lexicographer to compile such a lemmalist and (c) the compilation of the lemmalist is less reliant on cross-referencing as lexicographic device to combat decontextualisation brought about by strict alphabetical ordering.

The key consideration, however, for the compilation of an amalgamated lemmalist for this type of cognates is the degree/extent of the difference in form. Martin (2012b: 14) categorises such words as items with a small, systematic spelling or morphological difference or items with a bigger, non-systematic difference but which are still recognizably similar in form. He gives Dutch *pompoen* 'pumpkin' and *pinguïn* 'penguin' versus Afrikaans *pampoep* and *pikkewyn* respectively as examples. There is only a minor difference between *pompoen* and *pampoep* but a substantial difference between *pinguïn* and *pikkewyn*. For the Sotho languages a closer look at the degrees of similarity/difference in spelling is required and will be attempted in a hierarchical order from "very similar" to "more substantial" differences.

The first instance pertains to words which differ only in terms of a diacritic sign, e.g. *s* versus *š*. Setswana and Sesotho use the same word *ngwetsi* 'daughter-in-law' versus Sepedi *ngwetši*. Here the same letter (*s*) occurs — the only difference is *s* with or without the inverted circumflex "v" and there is no need to lemmatise *ngwetsi* and *ngwetši* as separate lemmas with *ngwetši* as a main lemma directly following *ngwetsi* in the vertical layout of alphabetical ordering. This paradigm could simply consist of a presentation indicating the names of all three languages, **ngwetsi**_[Set, Ses], **ngwetši**_[Sep] or an unmarked *ngwetsi* followed by a marked occurrence for *ngwetši*, i.e. **ngwetsi**, **ngwetši**_[Sep]. There is also no need for cross-referencing.

The second type of typical examples are words which differ in terms of a single letter. This single letter could be (a) different, or (b) added/omitted. Sepedi and Setswana has *ntatemogolo* 'grandfather' compared to Sesotho *ntatemoholo*

with "g" versus "h" as the only difference. As for *mogatsaka* versus *mogatšaka* such examples are less problematic — a single lemma paradigm will suffice e.g. **ntatemogolo**_[Sep, Set], **ntatemoholo**_[Ses]. As for (b), consider Sepedi and Setswana *morwa* versus Sesotho *mora* 'son'. The lexicographer, with the abilities of the target users in mind, must decide whether a user looking for *mora* will find it under *morwa*. In the case of 'daughter' in the Sotho languages both (a) and (b) apply, i.e. Sepedi *morwedi* differs from Setswana *morwadi* in respect of one letter and Sesotho *moradi* in respect of one omitted letter to Setswana. The difference between Sepedi and Sesotho however comprises both (a) and (b) and the question is whether the user looking for the lemma **moradi** will find it under **morwedi**. Given the alphabetical remoteness of "a" in *moradi* from "w" in *morwedi* (almost at opposite ends in an alphabetical stretch in the dictionary), it could be argued that they should both be lemmatised with cross-reference from the untreated lemma to the treated lemma(s).

Since an alphabetical ordering is followed, the degree of similarity or likelihood of recognition as cognates is influenced by the position inside the word where the differences occur, i.e. at the beginning, middle or end of the word. Spelling differences at the end or even in the middle of words are less problematic, e.g. *ntatemogolo* versus *ntatemoholo* but differences in the first few letters pose a greater risk of the user not finding the lemma e.g. *moradi* versus *morwedi* where the difference lies within the first four letters. Sepedi *ramogolo* and Setswana *rremogolo* also only differ in one instance of (a) and of (b) but although *ramogolo* and *rremogolo* is relatively easily recognisable as cognates when seen together, the user who wants to look up *rremogolo* will probably not see *ramogolo* because *ra-* is alphabetically remote from *rre-*.

The lemma paradigm to be considered often consists of more than one cognate for each of the languages. Consider for example the paradigm for *brother* or *sister* where Sepedi has *kgatšedi*, *dikgatšedi*, *bokgatšedi*, *kgaitšadi*, *bokgaitšadi*, *dikgaitšadi*, *kgaitšedi*, *bokgaitšedi*, *dikgaitšedi* versus Setswana *kgaitšadi*, *bokgaitšadi*, *dikgaitšadi*, *kgantsadi*, *bokgantsadi*, *dikgantsadi* versus Sesotho *kgaitšedi*, *bokgaitšedi*, *dikgaitšedi*. Assuming that the target users have a reasonable knowledge of at least one of the Sotho languages, and know the basic formation rule for singular/plural, these 18 forms could be reduced to *kgatšedi*, *kgaitšadi*, *kgaitšedi*, *kgaitšadi*, *kgantsadi* and *kgaitšedi*. This lemma paradigm could sensibly be further reduced on the basis of frequency: *kgatšedi* (635), *kgaitšedi* (39), *kgaitšadi* (4), *kgaitšadi* (90), *kgantsadi* (4), *kgaitšedi* (244) to *kgatšedi*, *kgaitšedi* _[Sep]; *kgaitšadi* _[Set]; *kgaitšedi* _[Ses].

For the lexicographer the extent of utilisation of cross-references is in the first place a measurable one. The norm followed in ANNA (Martin 2012b: 419) is that only members of a specific lemma paradigm which are alphabetically more than seven positions away from the lemma paradigm where treatment is given, must be cross-referred to the lemma paradigm. The number of such cross-references represents a redundancy factor against the success of the amalgamated approach because additional dictionary space is utilised for such

lemmas. Formulated differently, the more lemmas required to be entered separately from their lemma paradigms, the less successful the amalgamated approach will be because the ideal is to have a single lemma paradigm for each term for all three languages.

Cross-referencing is, however, intuitive in the sense of the presumed user's ability to find the lemma. In the case of *ntatemogolo* versus *ntatemoholo* it can be assumed that even the less knowledgeable user will be able to find the lemma but in cases such as *ramogolo* versus *rremogolo* the less sophisticated user should be assisted by including *rremogolo* as a lemma with cross reference to *ramogolo*.

5. Microstructural considerations

For short articles, e.g. consisting of little more than translation equivalents, such as for *mme*, *ntate* and *malome* given above, the success of an amalgamated approach is obvious. The question, however, is whether an amalgamated approach is still viable for longer articles.

Returning to the paradigm for *kgaetšedi* compiled above, consider the articles given for Sepedi (GNSW), Setswana (SESD) and Sesotho (SSED).

(GNSW)

KGÁETŠÉDI, (n-)/di- (*kgaetšedi*), cf. **KGÁITŠÁDI, KGÁITŠÉDI**, suster [so genoem deur broer], broer [so genoem deur suster] // sister [so called by brother], brother [so called by sister]; *lesogana lê sa êtego le nyala* ~ 'n mens verbreed jou kennis deur te reis // travel broadens the mind
KGÁETŠÉDI, -/bó- (*kgaetšedi*) v. **KGÁETŠÉDI, (n-)/di-**

(SESD)

kgaitsadi N. CLS. 1A₀- AND 9N-, same as *kgantsadi*, SING. OF *bokgaitsadi* and *dikgaitsadi*, sibling of the opposite sex; a sister; a brother.

(SSED)

kgaitsēdi (bọ̈) n.cl.1a, *kgaitsēdi* (dị) n.cl.5, brother of a woman, sister of a man; my brother, my sister; *kgaitse-die*, *kgaitsedi a hae*, his sister, her brother; *kgaitsedio*, your sister, your brother. |*kgàitsédi*|. *bokgaitsēdi*, n., state of being a brother or sister.

Consider the following attempts at the compilation of amalgamated articles for kinship terms:

kgaetšedi, kgaitšedi _[Sep]; **kgaitsadi** _[Set]; **kgaitsedi** _[Ses] (1a/2b/10) brother (so-called by sister), sister (so-called by brother) ♦ *lesogana le sa sepelego le nyala* ~ a young man who does not travel marries his sister: travel broadens the mind

The lemma paradigm in this example as well as its relatively short article has a high information density which can be paraphrased as follows. First, in terms of comment on form *kgaetšedi, kgaitšedi* _[Sep]; *kgaitsadi* _[Set]; *kgaitsedi* _[Ses] account for, compare and contrast, the most frequently used terms for all three of the Sesotho languages. This is indicated by the clear, functional and space saving convention _[Sep], _[Set] and _[Ses] in subscript. Secondly, noun class indication is given by a compact but clear convention. The boldfaced number indicates the class to which the lemmas belong, 2b and 10 the classes in which the plural forms occur. As for comment on semantics, the fact that the lemma can refer to a brother or a sister depending on the gender of the speaker is important and it is neatly explained by the brief contextual guidance given in brackets. Finally, the proverb given as an example of usage is well-selected because it is used in all three languages. The fact that the example is given in only one of the Sotho languages will not be problematic to the target user in this case because the forms are very similar in the other two languages. Thus no need to indicate the languages nor to attempt giving an example for each of the languages. Thus it saves dictionary space, also in terms of examples.

Consider also the suggested articles for *great grandfather* and *rakgolokhukhu*:

great grandfather rakgolokhukhu _[Sep], rremogolo/ntatemogolo_[Set], ntatemoholo_[Ses]

The treatment of the lemma **great grandfather** indicates that Sepedi uses the term *rakgolokhukhu* while Setswana uses *rremogolo* and *ntatemogolo* and that Sesotho also has the latter term with minor spelling variation, i.e. *ntatemoholo*.

rakgolokhukhu _[Sep] 1a/2b great grandfather. Also **mmelega rakgolo** who gave birth to grandfather. → **rremogolo/ntatemogolo**_[Set], **ntatemoholo**_[Ses]

This is an example where one of the three Sotho languages employs a unique term for a specific relationship while the other two use different terms. The user wants to find the meaning of *rakgolokhukhu* and looks it up under R in the dictionary. The treatment indicates that it is a Sepedi word in class 1a with plural form in class 2b and that the English translation equivalent is *great grandfather*. It also informs him/her of the alternative *mmelega rakgolo* and its literal meaning. Finally in the spirit of the amalgamated approach, i.e. to highlight similarities and differences, an explicit cross-reference by means of the reference marker "→" is given to the reference addresses for the Setswana and Sesotho terms **rremogolo/ntatemogolo**, **ntatemoholo** in the dictionary where more information can be found.

6. Using the convention for lemmatisation of kinship terms in an amalgamated approach

Prinsloo (2012) adapted the original *ga/sa/se* convention (Prinsloo and Gouws 1996) for the reduction of lemma paradigms for kinship terms in Sepedi. He indicated how a complicated set of derivations of *malome* 'uncle' such as *malomeago* 'your uncle', *malomeagwe* 'his/her uncle', *bomalomeabona* 'their uncles', etc. as well as a set of phrases involving *malome* could be reduced to a single lexicographic convention, i.e. **bo/mma/mogatša ~ ago/agwe**. In an amalgamated approach for the Sotho languages the question is whether *three* sets of complex derivations totalling more than 50 options could still be handled by a single convention taking the equivalents for *brother/sister* as a case in point.

Words related to *kgaetšedi* in Sepedi:

bodikgaetšedi	kgaetšedia	kgaetšediarena
bokgaetšedia	kgaetšediabona	kgaetšeditsebegokwa
bokgaetšediabo	kgaetšediago	ngwanakgaetšediago
bokgaetšediagwe	kgaetšediagogoba	ngwanakgaetšediake
bokgaetšediaka	kgaetšediagwe	polaokgaetšedi
dikgaetšedi	kgaetšediaka	ngwanakgaetšediago
kgaetšedi	kgaetšediake	sekgatšedi

Words related to *kgaitsadi* in Setswana:

bokgaitsadi	bokgaitsadiaka	dikgaitsadi
bokgaitsadia	bokgaitsadiake	kgaitsadi
bokgaitsadiabone	bokgaitsadie	kgaitsadia
bokgaitsadiagwe	bokgaitsadio	kgaitsadiarona
kgaitsadiabone	kgaitsadiake	kgaitsadio
kgaitsadiabona	kgaitsadialona	ngwanaakgaitsadiake
kgaitsadiaka	kgaitsadie	

Words related to *kgaitsedi* in Sesotho:

bokgaitsedie
dikgaitsedi
kgaitsedi
kgaitsedia
kgaitsediao
kgaitsedie
kgaitsedinyana
kgaitsedio
kgaitsediyaka
bokgaitsedi
bokgaitsediae
dikgaitsedie

It is simply not possible to lemmatise and treat all derivations of every single kinship term in the dictionary. It could be argued that, depending on the knowledge of the target user, it is possible to sensibly reduce these entire paradigms to a single convention for text reception purposes:

**bo/di/se/ kgaetsedi_[Sep] kgaityadi_[Set], kgaityadi_[Ses] /a/abone/abo/agwe/aka/ago/
arena/arona**

Such a convention requires detailed explanation in the users guide of the dictionary as has been done for the original *ga/sa/se* convention in POP.

7. Conclusion

The lemmatisation and treatment of kinship terms for a bi-directional dictionary bridging English and the Sotho languages in an amalgamated approach poses great challenges to the lexicographer on both the macro and microstructural levels.

On macrostructural level the first step will be to gather all single word basic terms, derived terms and phrases expressing kinship relations for all three Sotho languages and for English. The aim should be to compile a user friendly amalgamated lemmalist and that requires among other, insight and consideration of the presumed knowledge and dictionary using skills of the target user. Against this background of the user perspective the lexicographer should find a sound balance between the compilation of a lemma paradigm covering all three the Sotho languages versus separate lemmas, and utilisation of the medio structure. It is a matter of combating redundancy, i.e. to use less dictionary space for the lemmalist as long as user-friendliness in terms of the skills of the target user is not compromised. It has been argued in detail that the key consideration for the compilation of an amalgamated lemmalist is the degree/extent of the difference in form.

On the microstructural level the aim should be to achieve high text density which is still user-friendly and that clearly brings out differences and similarities between the amalgamated languages. Depending on the size of the dictionary more, or less comment on form and semantics could be given, i.e. longer or shorter articles as long as the information is well-balanced between the languages.

Where the reduction of huge quantities of terms, e.g. all derived forms of a specific term in all of the three Sotho languages is at stake, a lexicographic convention such as the adapted *ga/sa/se* convention could be used to combat redundancy and resolve the impossibility to lemmatise all the relevant forms. Care should, however, be taken that the compilation convention remains user-friendly, i.e. not attempting to include too many derivations.

The compilation of amalgamated dictionaries has great potential for African languages and the foundation laid by Martin's design and the publication

of ANNA is a source of inspiration to apply the model to closely related languages such as the Sotho and Nguni languages.

Acknowledgement

This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85763).

The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.

8. References

Dictionaries

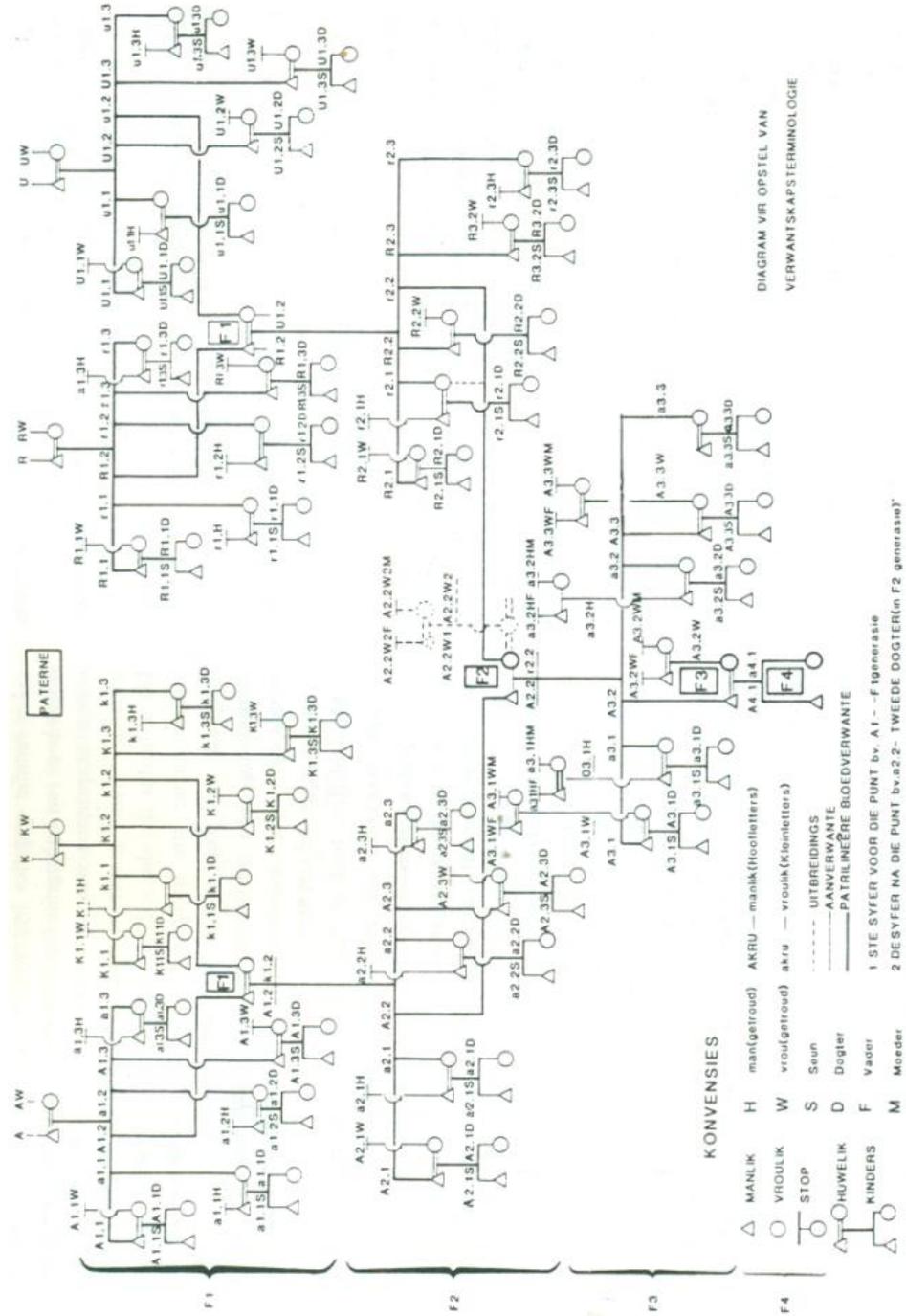
- ANNA = Martin, Willy (Ed.-in-chief). 2011. *Pharos Groot Woordeboek. Afrikaans en Nederlands (Prisma Groot Woordenboek Afrikaans en Nederlands)*. Cape Town: Pharos.
- BUKAN = Bukantswe <http://bukantswe.sesotho.org/>
- CSD = Dent, G.R. and C.L.S. Nyembezi. 1994. *Compact Setswana Dictionary English–Setswana, Setswana–English*. Pietermaritzburg: Shuter & Shooter.
- DSEA = Snyman, J.W., J.S. Shole and J.C. le Roux. 1990. *Dikišinare ya Setswana–English–Afrikaans Dictionary/Woordeboek*. Pretoria: Via Afrika.
- FREEL = <http://www.freelang.net/dictionary/sesotho.php>.
- GNSW = Ziervogel, D. and P.C. Mokgokong. 1975. *Pukuntšū ye kgolo ya Sesotho sa Leboa, Sesotho sa Leboa–Seburu/Seisimane/Groot Noord-Sotho-woordeboek, Noord-Sotho–Afrikaans/Engels/Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.
- LSS = Hlalele, J.B. 2005. *Longman Sethantšo sa Sesotho*. Maseru: Longman.
- NSSD = Chapole, S.R. 1997. *New South Sotho Dictionary*. Pietermaritzburg: Shuter & Shooter.
- ONSD = De Schryver, G.-M. et al. (Eds.). 2008. *Pukuntšū ya Polelopedi ya Sekolo. Sesotho sa Leboa le Seisimane. E gatišitšwe ke Oxford/Oxford Bilingual School Dictionary. Northern Sotho and English*. Cape Town: Oxford University Press.
- POP = Kriel, T.J., D.J. Prinsloo and B.P. Satheke. 1997. *Popular Northern Sotho Dictionary, Northern Sotho–English, English–Northern Sotho*. Cape Town: Pharos.
- PTLH = Mojela, M.V. (Ed.). 2007. *Pukuntšūtlhaloši ya Sesotho sa Leboa*. Pietermaritzburg: Nutrend.
- PUKU 1 = Kriel, T.J. 1983. *Pukuntšū woordeboek*. Pretoria: J.L. van Schaik.
- SESD = Matumo, Z.I. 1993 *Setswana–English–Setswana Dictionary*. Gaborone: Botswana Book Centre and Macmillan Botswana Publishers.
- SLEPD = Mojela, M.V., M.C. Mphahlele, M.P. Mogodi en M.R. Selokela. 2006. *Sesotho sa Leboa/ English Pukuntšū Dictionary*. Cape Town: Phumelela.

- SSED = Mabile, A. and H. Dieterlen. 1988. *Southern Sotho–English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.
- TYMYS = Otlogetswe, T.J. 2012. *Tlhalosi ya medi ya Setswana*. Gaborone: Medi Publishing.
- TYS = Mareme, G.B. (Ed.). 2008. *Thanodi ya Setswana: Sefala kgobokgobo*. Pietermaritzburg: Nutrend.

Other sources

- Bosman, Nerina.** 2013. Die gebruik van ANNA in 'n Nederlandse taalverwerwingskursus — toegangsgemak en inligtingskoste. Botha, W., P. Mavoungou en D. Nkomo. 2013. *Festschrift RUFUS H. GOUWS*: 39-54. Stellenbosch: SUN PRess.
- Marais, R.** 2011. Een mooie dikke dame. "ANNA", het Groot Woordenboek Afrikaans en Nederlands. *Ons Erfdeel* 1: 190-192.
- Martin, W.** 2012a. Amalgamated Bilingual Dictionaries. Genis, René et al. (Eds.). 2012. *Between East and West. Festschrift for Wim Honselaar on the Occasion of his 65th Birthday*. *Pegasus Oost-European Studies* 20: 437-449. Amsterdam: Uitgeverij Pegasus.
- Martin, W.** 2012b. ANNA: A Dictionary with a Name (and what Lies Behind it). *Lexikos* 22: 406-426.
- Martin, Willy and Rufus Gouws.** 2000. A New Dictionary Model for Closely Related Languages: The Dutch–Afrikaans Dictionary Project as a Case-in-point. Heid, Ulrich, Stefan Evert, Egbert Lehmann en Christian Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress. EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000*: 783-792. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Stuttgart University.
- Molalapata, B.T.** 2004. *The Treatment of Kinship Terminology in Sotho Dictionaries, with Special Reference to Setswana*. Unpublished M.A. Dissertation. Pretoria: University of Pretoria.
- Prinsloo, D.J.** 2005. *Compiling a Sotho–English/English–Sotho Dictionary: A Viability Study*. Unpublished paper presented at the Tenth International Conference of the African Association for Lexicography, held at the University of the Free State, Bloemfontein, South Africa, 13–15 July 2005.
- Prinsloo, D.J.** 2012 Die leksikografiese bewerking van verwantskapsterme in Sepedi. *Lexikos* 22: 272-289.
- 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 J.J. van Wyk.** 1992. Verwantskapsterminologie van die Noord-Sotho. *South African Journal of Ethnology* 15(2): 43-58.
- Van Wyk, J.J. and F.T. Haasbroek.** 1990. Verwantskapsterminologie van die Batswana. *South African Journal of Ethnology* 13(4): 159-179.

Appendix



The Inclusion of Word Formation in OALD8: The Case of Undefined Run-ons

Alenka Vrbinc (*alenka.vrbinc@ef.uni-lj.si*), Faculty of Economics,
and
Marjeta Vrbinc (*marjeta.vrbinc@ff.uni-lj.si*), Faculty of Arts
University of Ljubljana, Ljubljana, Slovenia

Abstract: The study presented in this contribution aims to investigate whether the print edition of OALD8 still sticks to the tradition of including many derivatives as run-on entries. For the purpose of the study, a database was compiled consisting of 1,200 lemmata with full entry status, and in this lemma range, 145 run-ons were found. The parts of speech of the lemma under which a run-on could be found were closely studied, and so were the illustrative examples. Nouns are most commonly included as run-ons, followed by adverbs, adjectives and verbs. The problem of polysemous entries at the end of which undefined run-ons can be found is also discussed. When it is difficult to draw parallels between the sense(s) of the lemma and the run-on deriving from it, it would be much better to include the run-on as the lemma and to define all its senses separately or to include undefined run-ons at the end of each individual sense of the lemma. Finally, it should be clear to the user when and why a certain word is included as a run-on and not as a lemma, and how s/he can establish a relation between the meaning of the lemma and the run-on.

Keywords: MONOLINGUAL LEARNER'S DICTIONARIES, OALD8, LEMMATA, SECONDARY LEMMATA, UNDEFINED RUN-ONS, PROPORTION OF RUN-ONS TO ENTRIES, EXAMPLES OF USE, DERIVATIVES, SUFFIXES, PART-OF-SPEECH LABELS, GRAMMATICAL INFORMATION

Opsomming: Die insluiting van woordvorming in die OALD8: Die geval van onverklaardes. Die studie waaroor daar in hierdie artikel berig word, probeer vasstel of die gedrukte weergawe van die OALD8 die tradisie handhaaf om baie afleidings as onverklaarde lemmata op te neem. Vir die doel van hierdie studie is 'n databasis saamgestel van 1,200 lemmata met volle artikels en tussen hierdie lemmata is 145 onverklaarde lemmata gevind. Die woordsoort van die lemmata waaraan onverklaardes geheg is, is noukeurig nagegaan asook die voorbeeldmateriaal. Naamwoorde tree die algemeenste as onverklaardes op, gevolg deur bywoorde, adjektiewe en werkwoorde. Die probleem van polisemiese inskrywings waaraan onverklaardes geheg word, word ook bespreek. Wanneer dit moeilik is om die verband tussen die betekenisonderskeidinge van 'n lemma en die voortspruitende onverklaardes te lê, sal dit beter wees om die onverklaardes as volle lemmata op te neem en die onderskeie betekenisonderskeidinge te verklaar, of om die onverklaardes aan die einde van elke verklaring van 'n betekenisonderskeiding te heg. Dit moet vir die gebruiker duidelik wees waar en wanneer 'n woord as onverklaarde opgeneem word en nie as volle lemma nie en ook hoe die verband tussen die betekenis van die lemma en die onverklaarde vasgestel kan word.

Slutelwoorde: EENTALIGE AANLEERDERWOORDEBOEK, GEBRUIKSVORBEELDE, GRAMMATIESE INLIGTING, LEMMATA, OALD8, ONVERKLAARDES, SEKONDÊRE LEMMATA, SUFFIKSE, VERHOUDING VAN ONVERKLAARDES TOT INSKRYWINGS, WOORDSOORTETIKETTE

1. Introduction

In the A–Z part of any dictionary, users can find lemmata listed in alphabetical order and accompanied by many pieces of information. Dictionary users definitely expect that a dictionary entry will present the facts as clearly as possible. The amount of information included in a particular dictionary depends a great deal on the type of dictionary; moreover, the needs and skills of potential target users based on the user profile should also be taken into account (Atkins and Rundell 2008: 200). Information about a lemma is especially detailed in monolingual learner's dictionaries, which are intended not only for decoding but also for encoding purposes. Learners of a foreign language need in-depth information that enables them to use a word correctly in context, which means that besides the meaning, they are also interested in pronunciation, grammatical information, grammatical patterns, collocations, synonyms, usage, etc. Apart from lemmata, two other types of lexical items within the entry should be mentioned: secondary lemmata and run-ons. What these have in common is that they both appear at the end of the entry in a special section that is often marked by a special symbol. However, the treatment of secondary lemmata differs from that of run-ons. Secondary lemmata, e.g., idioms and phrasal verbs, are defined and mostly also exemplified, whereas run-ons are not explained. What users learn about them is their part of speech; grammatical information is also provided, and sometimes their use is illustrated by means of one or more examples. Since the aim of this contribution is to study the inclusion and treatment of run-ons in OALD8, we will now take a look at some basic issues regarding run-ons.

Studying the existing literature shows that lexicographers are aware of a problem with the inclusion of run-ons. Stein (2002: 62), for example, points out that "the more the meaning of a combination is assumed to be inferable from the meaning of its constituents listed in the dictionary and the process of formation itself, the stronger the likelihood that it will not be listed as a dictionary item". Similarly, Béjoint (2010: 283) notices that "some items in the main A–Z list are treated as if they were of lesser importance, because they can easily be accessed via other words". Atkins and Rundell (2008: 237), on the other hand, emphasize the fact that run-ons can only be used in monolingual dictionaries when the word form is infrequent; when its meaning is unambiguously deducible through the application of basic word-formation rules; when its pronunciation can be predicted from the pronunciation of the headword to which it is attached; and when its grammatical and collocational behaviour is simple and predictable. Stein (2002: 50) also agrees that criteria should be determined

in order to decide under which conditions a certain derivative will be treated as a lemma and under which as a run-on entry. She mentions lexicalization as one of the criteria, but at the same time, she observes many inconsistencies when comparing existing dictionaries.

The most important question that should be posed is whether dictionary users can benefit at all from run-on entries. Stein (2002: 11) believes that this practice can somehow be justified in dictionaries intended for foreign learners because foreign learners are told that certain coinages exist in the language. On the other hand, she claims that native speakers would not really care whether or not derivatives ending in a specific suffix exist. Svensén (2009: 376) agrees with Stein in that nesting has a pedagogical purpose because bringing together lemmata that are morphologically related can promote word comprehension and vocabulary learning. It is, however, doubtful whether foreign learners could apply theoretical knowledge about word formation when looking up undefined run-ons. That is why Atkins and Rundell claim (2008: 235) that "run-on entries tend not to be used so much in learner's dictionaries, the idea being that learners have enough trouble finding what they want without having to burrow around in an entry of a headword that is not the object of their search". Also Stein (2002: 12) notices that some lexicographers and publishers are already trying to get away from run-on practice, the reason being that it is not user-friendly. According to Bergenholtz and Gouws (2013: 72), the use of a cumulative list of unexplained word formations with the lemma sign can be useful for the native speaker who only wants to check whether the word formation exists, how it is spelt or whether it is used with a linking morpheme or not.

Landau (2001: 102) believes that many dictionaries include even derivatives that may never have been used, simply to increase the number of entries in the dictionary. The selection of derivatives listed as run-ons is often arbitrary, thus making consultation of run-on entries difficult (Hartmann and James 1998: 121). Landau (2001: 102) points out that there are cases when the run-ons are more commonly used than the words to which they are run on, and the same observation is made by Atkins and Rundell (2008: 237). This means that lexicographers do not always regard the frequency of a certain word as the main criterion for inclusion of either the lemma or the run-on. Large linguistic corpora now enable lexicographers to determine the relative frequency of a particular word or of a particular sense, but according to Landau (2001: 104), relatively high frequency does not always guarantee lemma status, although it is obviously of great importance and is one of the most important factors to be considered. Atkins and Rundell (2008: 237) also draw attention to the fact that some derivatives are simply used too frequently to be handled as run-ons, and they may also show signs of unpredictable behaviour. Another problem the lexicographers encounter when including a certain word as a run-on is that a run-on may have acquired senses not adequately covered by the root word (Landau 2001: 102).

Traditionally, the main reason some words are included as run-ons is to conserve storage space, which is relatively restricted in traditional paper dic-

tionaries, as opposed to electronic dictionaries where it is relatively unrestricted, thanks to modern, high-capacity storage media (Lew 2011: 4). That is why the aim of our study is:

- to investigate whether the print edition of OALD8 still sticks to the tradition of including many derivatives as run-on entries,
- to gain insight into the proportion of run-ons to entries,
- to research what words merit the status of an undefined run-on and to which lemma the derivative is run on and, last but not least,
- to study additional information found in the run-on entries.

2. Methodology

In order to study the run-ons in OALD8, a database was compiled consisting of 1,200 lemmata, each with the status of a full entry. Six hundred lemmata were taken from a randomly chosen stretch in the letter C (i.e., from *caterer* to *chicken wire*) and six hundred from a randomly chosen stretch in the letter S (i.e., from *shwa* to *slip knot*). The material for the study was chosen on the basis of the following criterion: the letters C and S belong to two letters in English with the greatest number of words, thus according to Bukowska (2010: 1265) exhibiting greater variation than letters that have lower numbers of words. Apart from that, her recommendation is to choose a simple random selection of pages (ibid: 1267). Phraseological units (included in the idioms section) and multi-word verbs (included in the phrasal verbs section) that have the status of secondary lemmata were not included in the database, the only exception being those lemmata marked as verbs used in combination with particles only, which means that they appear only as a multi-word verb (e.g., *sick* is labelled as an adjective, noun and verb, but under the verb, only the multi-word verb *sick sth up* is entered and treated). All lemmata were typed into the database together with a part-of-speech label. If the entry contained an undefined run-on, the latter was also entered into the database together with a part-of-speech label. All run-ons in the database were then counted in order to determine the proportion of run-ons to entries. The total number of undefined run-ons in our database amounts to 145. The next step was to closely study the types of run-ons as regards their part of speech as well as the part of speech of the lemma under which a run-on could be found. Since the use of many undefined run-ons is illustrated by examples, these examples were also included in our database and were studied more closely to see whether they were sufficiently informative and intelligible to enable the user to better understand and use the word in question.

3. Results

The six hundred lemmata beginning with the letter C include 58 undefined

run-ons. The run-on *Central American* is counted twice, since it contains a multiple PoS label, i.e., adjective, noun. In the case of the 600 lemmata beginning with S, they include 87 undefined run-ons. Two run-ons, i.e., *sixteenth* and *sixtieth*, have multiple PoS labels, i.e., ordinal number, noun, and are therefore counted twice. If we compare the two randomly chosen stretches separately, it can be seen that there are far more undefined run-ons in the stretch with the letter S than in the stretch with C. However, since the database should be treated as a whole, it can be determined that the proportion of undefined run-ons to lemmata is 145 to 1,200, which means 1 : 8.3 and that 12.1% are indeed undefined run-ons. Table 1 shows the number of lemmata, the number of undefined run-ons by individual stretches and as a total and the proportion of lemmata to run-ons.

	Lemma	Undefined run-on	Proportion of lemmata to run-ons
Letter C	600	58	10.3 : 1
Letter S	600	87	6.9 : 1
Total	1,200	145	8.3 : 1

Table 1: The number of lemmata, the number of undefined run-ons and the proportion of lemmata to run-ons

Since 12.1% of all lexical items in our database constitute run-ons, it is now necessary to take a look at the parts of speech of the undefined run-ons. It can be seen that nouns are most commonly included as run-ons (i.e., 24 nouns beginning with C, 44 beginning with S, the total being 68 nouns or 46.9%). Nouns are followed by adverbs (i.e., 21 adverbs beginning with C and 26 beginning with S are included as run-ons, the total being 47 adverbs or 32.4%). Adjectives appear less frequently as undefined run-ons (i.e., 10 adjectives beginning with C and 12 beginning with S, the total being 22 adjectives or 15.2%), whereas verbs are rarely found as run-ons (i.e., 3 verbs beginning with each of the two letters, the total being 6 verbs or 4.1%). The only other part of speech that can be found as an undefined run-on in our database is ordinal numeral, which can be found only twice in the letter S (1.4%). Table 2 shows undefined run-ons by part of speech, by each letter and as the total number.

	Undefined run-ons by part of speech				
	Nouns	Adverbs	Adjectives	Verbs	Ordinal numerals
Letter C	24	21	10	3	/
Letter S	44	26	12	3	2
Total	68 (46.9%)	47 (32.4%)	22 (15.2%)	6 (4.1%)	2 (1.4%)

Table 2: Undefined run-ons by part of speech

It is necessary to take an in-depth look at the run-ons by parts of speech. As is evident from our database, the most frequent suffix attached to nouns entered as run-ons is *-ness* (20 nouns); this suffix is followed by *-ing* (13 nouns); the suffix *-er* is also quite common (11 nouns), whereas all other suffixes appear rarely (3 nouns ending in *-ity*, 2 nouns in *-th*, and only one noun each can be found ending in *-ization*, *-ability*, *-ment*, *-ist*, *-an*, *-cy* and *-ism*).

Apart from nouns derived by suffixes, our database also contains nominal undefined run-ons that can be regarded as cases of conversion, which means that they are homonymous with the lemma to which they are run on. Twelve nouns of this type were identified, e.g., *centigrade*, *centre-left*, *simper*, *simulcast*.

As regards the part of speech of the lemma to which nouns are run on, it can be seen that exactly half of all undefined nouns (i.e., 34 out of 68 nouns) appear at the end of the entry for an adjective with which they are semantically linked (e.g., *cautiousness* under *cautious*, *silliness* under *silly*); slightly less than a third (i.e., 20 nouns) are listed at the end of the entry for a verb (e.g., *centralization* under *centralize*, *signposting* under *signpost*), and slightly more than a sixth (i.e., 12 nouns) are enumerated at the end of the entry for another noun (e.g., *Catholicism* under *Catholic*, *sightseer* under *sightseeing*). Only two nouns are run on at the end of the entry for a numeral (i.e., *sixteenth* and *sixtieth*, in the entries for *sixteen* and *sixty*, respectively).

As regards the number of adverbial run-ons, they are second only to nouns. With very few exceptions (*centre stage*, *sidelong*, *sideward* (with its variant *sidewards*), *silky* and *sky-high*), the adverbs that are included in OALD8 and appear in our database end in *-ly*. An adverbial run-on that is included at the end of the entry for any other part of speech rather than under an adjective with which it is semantically linked can be regarded as the exception rather than the rule. Only four out of 47 adverbs in our database can be found in entries for other parts of speech: nouns (*centre stage*, *chauvinist*), verbs (*simper*) and other adverbs (*single-handed*). All other adverbs are included at the end of the entry for an adjective that can either be the root word used to form an adverb by adding *-ly* (e.g., *cautiously* is entered under the adjective *cautious*, *shyly* under *shy*, etc.) or a homonym (e.g., *sidelong*, *sky-high*).

As far as adjectives treated as undefined run-ons are concerned, it can be said that they are mostly included in the entries for the nouns to which they are semantically linked (18 out of 22 adjectives, e.g., *centralist* under the noun *centralism*, *slandorous* under *slander*), and very rarely under the verbs (only two adjectives, i.e., *cheering* under the verb *cheer* and *slighting* under the verb *slight*) or adverbs (only two adjectives, i.e., *sideways* and *single-handed* under their homonyms *sideways* and *single-handed*). The suffixes attached to adjectival undefined run-ons differ greatly, and in the majority of the adjectives in our database, only one example can be found per suffix. The suffixes are as follows: *-ic*, *-(e)d*, *-(i)an*, *-ist*, *-ing*, *-y*, *-al*, *-ous*, *-ly*, *-less*.

In our database, verbs rarely appear as undefined run-ons. There are only six verbs treated in this way, i.e., *catnap* (in the entry for the noun *catnap*), *caw*

(in the entry for the noun *caw*), *checkmate* (in the entry for the noun *checkmate*), *sideswipe* (in the entry for the noun *sideswipe*), *sightsee* (in the entry for the noun *sightseeing*) and *silk-screen* (in the entry for the noun *silk screen*). As can be seen, they are included in the entries for the nouns with which they are homonymous, the only exception being the verb *sightsee*.

Another interesting issue is that undefined run-ons are followed by different types of information. Some are only listed at the end of the entry for the lemma and are accompanied by a PoS label and (possibly) by grammatical information, whereas others also include examples illustrating the use of the run-on. There are 26 out of 58 run-ons beginning with the letter C with illustrative examples, and the same holds true of 31 out of 87 undefined run-ons beginning with the letter S.

	Undefined run-ons	Undefined run-ons + illustrative example(s)
Letter C	58	26
Letter S	87	31
Total	145	57

Table 3: Number of undefined run-ons and number of undefined run-ons with illustrative examples

4. Discussion

Since many words such as nouns ending in *-ance*, *-ence*, *-ency*, *-er*, *-ing*, *-ness*, *-ity*, *-ism*, *-tion*, etc., adjectives in *-able*, *-al*, *-ary*, *-ful*, etc. and adverbs ending in *-ly* (Svensén 2009: 132; Béjoint 2010: 284) can be defined by certain patterns that repeat (e.g., *-ness* is supposed to mean "the state of being ...", *-ly* is supposed to mean "in a ... manner", etc.) (cf. also Landau 2001: 102), space in print dictionaries is often saved by including such derivatives at the end of the entries for the related lemmata as undefined run-ons. Space saving is no longer an issue in electronic or online dictionaries, but the subject of our investigation is a print dictionary, where space is still a factor that should be taken into account. However, if the undefined run-ons found in the print edition are checked in the online version of OALD as well as the CD-ROM version accompanying the print edition, it can be seen that the same lexical items are included as run-ons as in the print edition, even though there are no space limitations in electronic dictionaries. The only difference is that the online version and the CD-ROM version include illustrative examples in some run-ons which are not exemplified in the print edition (e.g., *shyness* and *shyly*). Dictionary users do not concern themselves with problems of space or other limitations, but they may be faced with difficulties when performing a look-up operation, since they most certainly expect that the word they seek will be included and treated as a lemma and not as a run-on. A lemma is easy to find because it is part of the

dictionary macrostructure, it is precisely defined, grammatical codes and grammatical patterns are provided, the use of the lemma is exemplified, collocations are included, etc. On the other hand, an undefined run-on is much more difficult to find, as it is hidden within the dictionary microstructure, and far fewer pieces of information about it can be obtained. (These considerations apply to print dictionaries only, since finding a lexical item in an electronic dictionary poses no problems at all.) Apart from that, it is implicitly assumed that users are capable of an immediate constituent analysis (Stein 2002: 54) when they come across a derivative listed as a run-on. What may be particularly troublesome is the absence of definitions, since dictionary users are mostly interested in the meaning of a word when they look it up (cf. Atkins and Rundell 2008: 405). Another issue that should be addressed is the criteria for including a certain word either as a lemma or as a run-on. Doubtlessly, dictionary users cannot possibly be expected to know how the lexicographers determine the status of a particular lexical item, which means they do not know where to look up a word — in the macro- or microstructure.

In the print edition of OALD8, the policy of including run-ons is briefly mentioned in the *Key to dictionary entries* (p. ix):

Some words that are **derivatives** of other words do not have their own entry in the dictionary because they can be easily understood from the meaning of the word from which they are derived (the root word). They are given in the same entry as the root word, in a specially marked section.

Users of the CD-ROM and the online versions, however, are completely deprived of this information. Nowadays, in the age of modern technology, it would be over-optimistic to assume that the majority of dictionary users would consult a print edition of any dictionary, since the advantages of electronic versions are so numerous that they far outweigh the benefits of print dictionaries.

As can be seen in the Results section, many nominal run-ons are characterized by typical suffixes. It is, however, questionable whether an average dictionary user is able to grasp the meaning of the derivative, since it does not suffice to know the meaning of the lemma from which the undefined run-on is derived — the user also needs to be familiar with the meaning of the suffix. For example, if one looks up the suffix *-ing* in OALD8, one learns that it is 'used to make the present participle of regular verbs'; this definition does not give any semantic information but rather informs the user about the grammatical characteristics of the suffix. The suffix *-ization* is an undefined run-on in the entry for the suffix *-ize*, whose meaning is divided into three senses: 'to become, make or make like', 'to speak, think, act, treat, etc. in the way mentioned' and 'to place in'. The suffix *-er* when used to form nouns has four senses in OALD8: 'a person or thing that', 'a person or thing has the thing or quality mentioned', 'a person concerned with' and 'a person belonging to'. Were the user to be somehow familiar with these meanings, s/he might not find it too difficult to infer the meaning of, for example, *chain-smoker* ('a person that smokes cigarettes

continuously'), after having studied the meaning of the verb *chain-smoke*, which is defined as 'to smoke cigarettes continuously, lighting the next one from the one you have just smoked'. It is, however, highly unrealistic to expect the user of a monolingual learner's dictionary to be aware that a certain suffix may also convey a certain meaning. Even if the user did know that, it cannot be expected that s/he would be familiar with the meanings of so many different suffixes, especially if we take into account that many suffixes have different senses. It could perhaps be anticipated that the suffix *-ability* would pose fewer problems, but this may be at least partly because, as far as its form is concerned, it is identical with the noun *ability*; consequently, the user who is unfamiliar with this noun can look it up and learn the meaning. Interestingly, the suffix *-ability* and its variant *-ibility* form an entry on their own, but the user is cross-referred to the entry for the suffix *-able*, which is, together with its variant form *-ible*, defined as 'that can or must be' and 'having the quality of'. At the end of the entry for the suffix *-able/-ible*, two undefined run-ons can be found, i.e., *-ability*, *-ibility* and *-ably*, *-ibly*, which means that the suffix *-ability/-ibility* can actually be found in two different places within one dictionary, but without being explained. To sum up, it would be advisable to provide definitions for such nouns instead of just running them on, because the meaning of the noun in question would be clearer and more easily understood by the majority of users.

As has been mentioned in the Results section, nouns with the status of an undefined run-on can also be appended to the homonymous lemma. The user who looks up, for example, *cerise* would learn that this adjective means 'pinkish-red in colour'. Hopefully, s/he would be able to decode the meaning of the nominal undefined run-on *cerise* as 'a pinkish-red colour'. Similarly, the verb *simper* is defined as 'to smile in a silly and annoying way', which means that users should somehow "guess" that the undefined noun *simper* means 'a silly smile'. This may seem an easy task, but many users would most certainly find it daunting, if not impossible.

It is well-known that learner's dictionaries tend to include precise, complete information about the grammar of the lemma. If nominal run-ons are studied more carefully, we could see that dictionary users can also learn something about the grammatical behaviour of these words. The same policy as for the defined lemmata is also employed for the run-ons. If the grammar code in nouns is absent, it means that the noun is countable (e.g., *signer*, *sight-seer*, etc.). Some nouns are followed by examples of use which sometimes make it clear that the noun is countable (e.g., both examples illustrating the use of the noun *signer*, i.e., *the signers of the petition*; *signers communicating information to deaf people*). Run-ons that are uncountable or always used in the singular are equipped with grammar codes: [U] (e.g., *centralization*, *centre-left*, *sincerity*, *sing*, etc.) or [singular] (e.g., *sizzle*). Again, examples of use, if provided, additionally illustrate the grammatical characteristics. At first sight, the policy of providing grammatical information about undefined run-ons seems not to be consistently pursued. The question can be raised whether the user is supposed to know that

countable nouns are not labelled, and another question that can be asked is how and/or where the user can learn this. In the section entitled *Key to dictionary entries* (pp. ix-xi), no explanation is provided, but the user is referred to pages R11-12 in the back matter, where the following information on different types of nouns can be found:

Countable nouns are the most common type of noun. If they have only one meaning, or if all the meanings are countable, they are just marked *noun*. For nouns that have a number of meanings, some of which are countable, each meaning that is countable is marked [C].

It seems overly optimistic to expect an average dictionary user to read the *Reference Section* in order to be able to learn all the details regarding the (grammatical) labelling of lemmata and run-ons. Would it not be more practical and user-friendly to assign each noun a grammatical code? In this case, the user would not have to study the front or back matter to find information about the inclusion or omission of grammatical information. On the other hand, the grammatical labelling of nouns is very detailed, since a distinction is made between uncountable nouns [U] and singular nouns [singular]. Again, the question can be posed whether average dictionary users will be able to understand and tell the difference between both types of nouns, even if they have read the explanation on page R11:

Uncountable nouns [U]: "An uncountable noun has only one form, not a separate singular and plural. It can be used with or without a determiner." [...] "With nouns such as **furniture**, **information** and **equipment**, as with many other uncountable nouns, you can talk about amounts of the thing or separate parts of the thing by using phrases like **a piece of**, **three items of**, **some bits of**. Nouns like **piece**, **item** and **bit** are called PARTITIVES when used in this way."

Singular nouns [singular]: "Some nouns are always singular and have no plural form. Many nouns like this can be used in only a limited number of ways. For example, some singular nouns must be or are often used with a particular determiner in front of them or with a particular preposition after them. The correct determiner or preposition is shown before the definition."

To sum up the discussion about the grammatical labelling of nouns, it would be advisable to use a suitable code to label each noun, regardless of its status (lemma or run-on) and its grammatical characteristics.

Adverbs, especially those formed by adding *-ly* to the related adjective, are one of the commonest types of run-ons (Atkins and Rundell 2008: 237), which is also in line with the results obtained when studying our database (for a more detailed discussion of adverbial run-ons in British monolingual learner's dictionaries cf. Vrbinc and Vrbinc 2013). Atkins and Rundell (2008: 237) warn that care needs to be taken when some adverbs in *-ly* are used as intensifiers (no such example can be found in our database), or when some adverbs have a double function: manner adverbs (e.g., *reply in characteristically robust style*) and sentence adverbs (e.g., *Characteristically, Helen paid for everyone*). They recom-

mend that in learner's dictionaries, it is best to avoid anything other than simple manner adverbs in the run-on slot, since such adverbs would require two separate senses: one sense for the manner adverb and one for the sentence adverb. As an example of such a treatment that would also be advisable in monolingual learner's dictionaries, we can take the two definitions for *characteristically* from *Collins English Dictionary* online:

1. in accordance with the nature of the person or thing involved, typically ⇒ *Hayes was characteristically modest about his achievement.* ⇒ *He replied in characteristically robust style.* ⇒ *a characteristically polished performance*
2. as was characteristic of him/her/it; predictably ⇒ *Characteristically, Puccini was dissatisfied with the end result.* ⇒ *Characteristically, he worked hard at the assignment.* ⇒ *She knew she was indeed dying; characteristically she did not tell her fellow actors.*

The third most frequent part of speech among the run-ons included in our database is the adjective. The users are doubtlessly faced with a dilemma in decoding the meaning of the run-on adjective because they are supposed to be familiar with the meaning(s) of individual suffixes. The suffix *-less*, for example, could perhaps be considered more transparent as regards its meaning ('without') than suffixes such as *-ic* ('connected with'), *-y* ('full of; having the quality of'), *-ous* ('having the nature or quality of'), etc. It is, however, not to be expected that the average users will first recognize the suffix, then look up the suffix in the dictionary and study its meaning and finally, decode the meaning of the adjective with a particular suffix. For this reason, it would be advisable to define such adjectives rather than just run them on. For example, *slanderous* can be found under the noun *slander*, which is defined as 'a false spoken statement intended to damage the good opinion people have of sb; the legal offence of making this kind of statement', which means that the user is supposed to decipher the meaning of the adjective *slanderous* from the definition for the noun *slander* and the definition for the suffix *-ous*. It would, however, be more user-friendly to define *slanderous* as 'untrue and intended to damage the reputation of the person that it refers to', thus enabling the user to get comprehensive information in one place rather than looking up the noun and the suffix and trying to construe the meaning.

Although included as undefined run-ons, adjectives may also have a grammatical label in order to show users how to use them correctly (cf. also *Reference Section* in the back matter of OALD8, p. R12). For example, the adjective *slighting* has the label [only before noun], which clearly tells users that the adjective is only used attributively.

Verbs are only exceptionally included as run-on entries — only six were found in the stretches studied for the purpose of our research. The question that should be addressed is why these verbs are not awarded full entry status. The answer may be sought in the frequency of these verbs. The compilers of

modern dictionaries, including monolingual learner's dictionaries, claim that dictionaries rely on various corpora. OALD8 is no exception, and on the back cover, we learn that it is "based on the authority of the British National Corpus and the Oxford English Corpus". It is generally known that a well-balanced corpus shows a great variety of contexts, thus providing a good source of definitions for most words. Apart from that, a corpus offers numerous examples of a particular usage and supplies the compilers with information about the frequency of the word in question (cf. also Landau 2001: 296-297). Does this mean that by measuring relative frequency, the corpus told the compilers of OALD8 that the above-mentioned verbs were not worth being treated as lemmata? If we check the frequency of verbs included as run-ons and compare it with the frequency of the nouns to which the verbs are run on, we get the following results:

	BNC (number of hits)		ukWaC (number of hits)	
	noun	verb	noun	verb
catnap	4	5	31	33
caw	3	2	121	22
checkmate	15	13	159	71
sideswipe	8	2	95	26
sightseeing	184	/	4,073	/
sightsee	/	14	/	97
silk screen	0	0	0	0
silk-screen	2	0	53	0

Table 4: Number of hits in BNC and ukWaC

As is evident from Table 4, frequency was clearly the criterion, since only in *catnap* is the number of hits for the verb slightly higher than for the noun. In all other cases, the number of hits for the noun exceeds that for the verb. It is, however, true that in *catnap*, the number of hits in both corpora is practically the same for the noun as for the verb, and the same holds true of the number of hits for *caw* and *checkmate* in the BNC. In the ukWaC, the number of hits for the noun and the verb *caw* and *checkmate* is distinctly different. A great difference in the number of hits in both corpora can also be observed in the case of *sideswipe* as well as *sightseeing* vs. *sightsee*. As far as the spelling of *silk-screen* is concerned, it can be seen that both corpora include the hyphenated spelling only, and they both agree that this lexical item is a noun and not a verb. Although frequency seems to play an important role, we should be aware of the fact that a corpus is a source of evidence rather than a source for every single decision made by a lexicographer. A corpus is essential in guiding the compiler to make important decisions, but eventually it is the compiler who needs to take decisions that satisfy the needs of potential dictionary users to the greatest extent possible.

As far as the provision of grammatical information for verbs is concerned, users are supposed to study the *Reference Section* in the back matter of the dic-

tionary (pp. R5-R8). In this section, they learn that the codes [T] and [I] are used to refer to transitive and intransitive verbs, but if a verb is always transitive in all its meanings, it is just marked *verb*. Here, we can draw a parallel with the labelling of nouns, where countable nouns are also not labelled if they are countable in all senses. However, it would be advisable for the sake of consistency to label all the verbs [T] and [I], irrespective of whether they may be transitive in all senses. This does not hold true just of verbs treated as entries but also of verbs included as run-ons. Sometimes other labels are added to draw users' attention to a certain peculiarity, e.g., besides including the grammatical code [intransitive], the verb *sightsee* also includes the label (only used in the progressive tenses). Since transitive verbs can take different types of object, the different patterns or verb frames are shown in bold type, usually just before an example showing that pattern in context, e.g., the verbs *checkmate* and *sideswipe*, which can be used in the pattern ~ **sb/sth**, or the verb *silk-screen* in the pattern ~ **sth**.

As has been established in the Results section, the examples of use are provided for 57 out of 145 undefined run-ons, i.e., 39.3%. Generally speaking, the basic criteria that should be satisfied if an example is to be described as a good example are naturalness and typicality, informativeness and intelligibility, and the right balance between these three criteria is of great importance if the examples of use included in the dictionary are to bring real benefits for the users (Atkins and Rundell 2008: 458-461). We should now examine the run-ons whose use is exemplified to see whether the examples meet these criteria and whether they help the reader either to better understand the meaning of the run-on or to use the run-on correctly.

Doubtlessly, an example should clearly show the context in which a lemma or a particular sense of the lemma typically occurs, which also holds true for examples illustrating the meaning(s) of run-ons. These examples should be chosen with the utmost care, since run-ons lack definitions, and users can only rely on examples if these have been provided. Even words that seem not to pose problems, being regarded as straightforward, may, in fact, not be so straightforward to a foreign learner and dictionary user who may face challenges ranging from problems with grammar to problems with collocations. From this point of view, some examples of use seem not to meet these criteria:

side-splittingly funny (used to illustrate the use of the undefined adverb *side-splittingly*);

a sideways move (used to illustrate the use of the undefined adjective *sideways*);

a vow of celibacy (used to illustrate the use of the undefined noun *celibacy*);

ceramic tiles (used to illustrate the use of the undefined adjective *ceramic*).

Because these examples are used in truncated form and lack context, users do not get any information whatsoever about how or in what kind of context to

use the run-ons correctly. Also, the examples that are supposed to illustrate the meaning of agent nouns probably do not make much sense, since they contribute neither to the understanding nor to the correct use of the words in question. For example:

He's an Olympic silver medallist (used to illustrate the use of the undefined noun *silver medallist*).

Since users aim to construct English sentences that are syntactically, colligationally and collocationally natural, the illustrative examples should incorporate as many pieces of information as possible. Besides that, it should be clear from the illustrative examples what the word defined really means. This can be illustrated by the following examples taken from our database:

The game will be broadcast simultaneously on TV and radio (used to illustrate the use of the undefined adverb *simultaneously*);

The hotel is centrally located for all major attractions (used to illustrate the use of the undefined adverb *centrally*).

In some examples, part of the example is printed in bold, which should immediately catch the users' attention. The bolded part of the example indicates a collocation (cf. also p. R13 in the *Reference Section* of OALD8, where it is explained that important collocations are printed in bold type within the examples). For example:

*a **cheerleading** squad/team* (used to illustrate the use of the undefined noun *cheerleading*);

*I can say **in all sincerity** that I knew nothing of these plans* (used to illustrate the use of the undefined noun *sincerity*).

Here, a parallel can be drawn between the examples for *cheerleading* and the truncated examples mentioned above. Although collocations are useful for a foreign learner, it would be more sensible to contextualize them rather than list bare collocations only.

If the compilers of OALD8 believe that the meaning of the example is not obvious, a short explanation follows in brackets. Some examples of this kind can also be found in our database:

*Is the house **centrally heated*** (= does it have central heating) (used to illustrate the use of the undefined adverb *centrally*);

*I could **cheerfully** have killed him when he said that* (= I would have liked to);
*She **cheerfully** admitted that she had no experience at all* (= she wasn't afraid to do so) (both examples illustrate the use of the undefined adverb *cheerfully*).

As Atkins and Rundell (2008: 460-461) state, the information in the example

should not be in conflict with what the definition says. The adjective *cheerful* is defined as '1. happy, and showing it by the way that you behave 2. giving you a feeling of happiness'. From this point of view, the example *I could cheerfully have killed him when he said that* seems to contradict the definitions, because *cheerfully*, which is supposed to imply something positive, is used with *kill*, which has a negative connotation. Atkins and Rundell (ibid: 461) rightly point out that such an example would be "a perfectly natural thing to say in the right context, but for a learner who has struggled to process the definition, and who believes s/he has grasped the concept, it can only be discouraging to find an example that seems to contradict all this".

As has been mentioned, intelligibility is one of the criteria required to make a good example. This means that, although an example is authentic, natural and typical, it may fail to be informative because the user is unable to understand it. Consequently, such an example is of no use or value to the user (Atkins and Rundell 2008: 461). An example taken from our database can best illustrate this:

Cheerleading used to be peachy blondes waving pompoms at football games (used to illustrate the use of the undefined noun *cheerleading*).

The words that many users would probably find incomprehensible are *peachy* (yielding 23 hits in the BNC, but these hits refer mostly to sense 1 in OALD8, i.e., 'like a peach in colour or appearance', and not to sense 2, i.e., '(NAmE, informal) fine; very nice') and *pompom* (yielding only 2 hits in the BNC). Apart from that, this example is also sexist and dependent on outdated cultural information.

Another example that is worth mentioning is the undefined noun *skewbald*, which is exemplified by the sentence *He was riding a skewbald*. It is questionable whether the dictionary user will be able to decode the meaning of this noun by simply reading the definition of the associated adjective *skewbald*, i.e., '(of a horse) with areas on it of white and another colour, usually not black'. It can be claimed with a high degree of certainty that the majority of users will not pay particular attention to the sense indicator provided in brackets, thus failing to understand that the noun *skewbald* actually means 'a horse or pony with patches of white and another colour, but usually not black'.

Another issue that should be addressed is the problem of a polysemous entry at the end of which an undefined run-on can be found whose use is illustrated by means of an example of use or several examples of use. In OALD8, the adjective *central* has five senses: 'most important', 'having power or control over other parts', 'in the centre of an area or object', 'easily reached from many areas', '(of a vowel) produced with the centre of the tongue in a higher position than the front or the back'. The adverb *centrally* is an undefined run-on, whose use is illustrated by three examples, i.e., *The hotel is centrally located for all major attractions*, *a centrally planned economy* and *Is the house centrally heated* (= does it have central heating)? It is evident from these three examples that the

meaning of *centrally* can be surmised in the first two examples if the user is familiar with the meanings of the related adjective but not in the third example, where *centrally* is explained by 'does it have central heating'. In this example, the dictionary user needs to know the meaning of the compound 'central heating' (which is an entry in its own right in OALD8), rather than any of the meanings of the adjective *central*. Since the derived adverb is placed at the end of the entry for the root word, which may have more than one sense, another question should be raised, i.e., to which of the root word's senses does the derived adverb belong? Does it belong to all of them or just to one of them? In some cases, it may be easier to connect the individual examples illustrating the use of run-ons with a particular sense of the lemma at the end of which the run-on is listed. This is mostly the case if a polysemous lemma has fewer senses, and if the undefined run-on contains examples illustrating the use of all senses of the lemma. For instance, let us consider the two examples found under the undefined verb *checkmate*. It can be seen that the example *His king had been checkmated* is linked to sense 1 of the noun *checkmate*, i.e., '(in chess) a position in which one player cannot prevent his or her king (= the most important piece) being captured and therefore loses the game', whereas the example *She hoped the plan would checkmate her opponents* is linked to sense 2 of the noun *checkmate*, i.e., 'a situation in which sb has been completely defeated'.

Quite rightly, the question should be raised how anyone can expect an average dictionary user to know whether the meaning of a particular run-on as suggested in the examples is semantically derived from the first, second, third or other definition of the lemma, or perhaps from all of them. Since OALD8 is intended for learners of English, such treatment of undefined run-ons should be considered undesirable because it is not to be expected that non-native speakers of English would be able to draw any parallels between the meaning(s) of the lemma and the run-on that is derived from it. In all such cases, it would be much better to include the run-on as the lemma and define all its senses separately. If, on the other hand, compilers of a dictionary have to include as many items as possible and if they have to save space at the same time, they necessarily resort to undefined run-on entries. This being the case, the following should be observed: only those words can be included as undefined run-ons whose meanings are truly transparent and easily inferred from the meanings of root words. This means that derived words that have acquired senses not adequately covered by the root words to which they are run on should be included only as main entries (e.g., *centrally heated*). Another solution to the problem is a practice employed by COBUILD7. This dictionary treats undefined run-ons in a more user-friendly way, since they are listed at the end of each individual sense of the lemma, together with at least one example of use. Such a treatment makes it easier for the user to establish a connection between the meaning of the lemma and the run-on.

As regards the inclusion of run-ons, one of the most obvious problems users may face is disruption of correct alphabetical order. For example, *cheerily*

is included under the adjective *cheery*, *silkinness* and *silkiy* under the adjective *silky*, so that alphabetical order is clearly disrupted. Strict alphabetical order is breached by the nesting of derivatives at the end of a certain entry, a practice which may not be transparent to many users (Hartmann 2001: 64). Such treatment may lead to more or less serious difficulties with the look-up process in print dictionaries (the electronic versions do not pose such problems). It should, however, be stressed that the success of any look-up operation may lie in the degree of the user's ability and knowledge about how to retrieve different pieces of information from a dictionary.

The last remark that should be made regarding the material collected for the purpose of our study concerns the undefined noun *sledding*, which is run on to *sled*. *Sled* is labelled with a multiple PoS label, i.e., noun, verb, and has a cross reference to *sledge*. This means that the user cannot possibly decipher the meaning of *sledding* because s/he cannot find any definition for *sled*. In the entries for the noun and the verb *sledge*, *sledding* is not mentioned. The best solution would probably be to include *sledding* as a lemma and provide a definition for it. Another such example is the undefined noun *Catholicism*, with a cross reference to *Roman Catholicism*, which appears at the end of the entry for the noun *Catholic*, where the user is cross-referred to *Roman Catholic*. In cases where a user is cross-referred from one entry to another without being able to find any definition for the particular word, the question can be asked what is actually the point of including such run-ons in a dictionary. It should, however, not be forgotten that OALD8 belongs to the category of dictionaries intended for foreign learners; thus, the inclusion and treatment of various types of information should be carefully considered to make them as user-friendly as possible.

5. Conclusion

Theoretically, the meaning of undefined run-ons should be decoded in two ways: with the help of the definitions for the lemma from which the run-on is derived, and with the help of a combination of these definitions and the definitions for the suffix (if the suffix has a semantic meaning). These two methods are problematic in several respects. Firstly, the polysemous lemma and its run-on do not necessarily coincide in all the senses; consequently, the user cannot be supposed to know which senses of the lemma s/he can apply to the run-on. Secondly, the semantic meaning of the suffix is very vague; thus, the definitions are difficult to understand, let alone to apply to the combination base + suffix. Most importantly, we should not forget that OALD8 belongs to the group of learner's dictionaries, which means that the users' native language is not English. This means that the entry should be tailored in a way that is most suitable for the intended users, that is foreign learners. Even if lexicographers provide one or more examples illustrating the use and/or context of the run-on, this does not necessarily help the user to decode the meaning of the run-on.

These problems can be solved in two ways: either by listing the undefined run-ons at the end of each individual sense of the lemma, together with at least one example of use (this method is employed by COBUILD7), or by including and treating polysemous derivatives as entries in their own right and not as run-ons. Finally, it should be clear to the user when and why a certain word is included as a run-on and not as a lemma, and how s/he can establish a relation between the meaning of the lemma and a run-on that is listed at the end of the entry for the lemma.

In print dictionaries, space saving has been a convenient "excuse" for various more or less user-friendly practices for decades — one of these being the inclusion of run-ons, whose main aim was to achieve broader coverage at a low cost in terms of space. This has become rather outdated in the era of modern technology, when even lexicographers often express doubts about the future of print dictionaries (Macmillan, for example, announced in November 2012 that, in future, only online dictionaries will be available), and if print dictionaries are a thing of the past, space-saving methods need no longer be employed. However, it should be emphasized that print dictionaries are still an entirely dominant force in many developing countries and will probably remain so for many years to come.

To conclude, we should bear in mind that dictionary users who look up an unknown or unfamiliar word in a dictionary expect to find the meaning of this word. If the word is included as an undefined run-on, they may still fail to understand it and may consequently be unable to use it. This is also in line with the observations made by Bergenholtz and Gouws (2013: 73, 75), who rightly speak in favour of a presentation and treatment determined by the users' needs.

Bibliography

Dictionaries

Carroll, Katherine (Ed.). 2012⁷. *Collins COBUILD Advanced Dictionary of English*. Glasgow: HarperCollins. (COBUILD7)

Collins English Dictionary. Available at <http://www.collinsdictionary.com/dictionary/english>.

Turnbull, Joanna (Ed.). 2010⁸. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press. (OALD8)

Other literature

Atkins, B.T. Sue and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Béjoint, Henri. 2010. *The Lexicography of English*. Oxford: Oxford University Press.

Bergenholtz, Henning and Rufus H. Gouws. 2013. The Presentation of Word Formation in General Monolingual Dictionaries. *Lexikos* 23: 59-76.

- Bukowska, Agnieszka Anuszk.** 2010. Sampling Techniques in Metalexigraphic Research. Dykstra, Anne and Tanneke Schoonheim (Eds.). 2010. *Proceedings of the XIV EURALEX International Congress*: 1258-1269. Ljouwert: Afûk.
- Hartmann, Reinhard R.K.** 2001. *Teaching and Researching Lexicography*. Harlow: Pearson Education.
- Hartmann, Reinhard R.K. and Gregory James.** 1998. *Dictionary of Lexicography*. London/New York: Routledge.
- <http://www.macmillandictionaryblog.com/bye-print-dictionary>
- Landau, Sidney I.** 2001². *Dictionaries: The Art and Craft of Lexicography*. New York/Cambridge: Cambridge University Press.
- Lew, R.** 2011. Space Restrictions in Paper and Electronic Dictionaries and Their Implications for the Design of Production Dictionaries. Bański, Piotr and Beata Wójtowicz (Eds.). 2011. *Issues in Modern Lexicography*. München: Lincom Europa. Available at <https://repozytorium.amu.edu.pl/jspui/bitstream/10593/799/1/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf>. [Accessed 15 May 2014.]
- Stein, Gabriele.** 2002. *Better Words: Evaluating Efl Dictionaries*. Exeter: University of Exeter Press.
- Svensén, Bo.** 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge: Cambridge University Press.
- Vrbinc, Alenka and Marjeta Vrbinc.** 2013. The Status of Adverbs in British Monolingual Learner's Dictionaries: Lemmata or Run-ons. *Poznań Studies in Contemporary Linguistics* 49(2): 261-274.

Corpora

British National Corpus

https://ca.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/bnc2;

ukWaC https://ca.sketchengine.co.uk/bonito/run.cgi/first_form?corpname=preloaded/ukwac3;

La "main invisible" dans les langues et la confection de dictionnaires locaux

Edgard Maillard Ella, *Institut de Recherches en Sciences Humaines, Centre National de la Recherche Scientifique et Technologique, Libreville, Gabon (ellamaillard@yahoo.fr)*

Résumé: Loin d'être leurs intentions, il nous semble que c'est comme mus par une main invisible, semblable à celle qui a été développée en économie par Adam Smith, que les locuteurs, dans la recherche de la performance individuelle, soient conduits à contribuer à l'enrichissement de leurs langues sans que cela soit leur intention. Pour Adam Smith, c'est en recherchant l'enrichissement personnel que les individus enrichissent aussi la nation. Pour lui, c'est une main invisible qui les conduit à atteindre un objectif qui n'était pas dans leurs intentions, d'autant plus que c'est dans ce cas qu'ils sont bien plus rentables à la société que s'ils avaient l'intention d'y contribuer. C'est fort de ces similarités entre les faits de la main invisible en économie, et ce qui semble être les manifestations des mêmes faits dans les langues, que nous tentons d'y suggérer l'énonciation de la même théorie. Utiliser une théorie économique pour expliquer certains faits de langues n'est pas nouvelle. Le sociolinguiste français Louis-Jean Calvet a trouvé une similarité entre la loi du marché, qui détermine la valeur d'échange d'une monnaie, et les valeurs respectives des différentes langues du monde qui sont soumises à leur degré de diffusion. En nous appuyant sur la théorie de la main invisible en économie, nous discuterons d'abord de l'énonciation de cette même théorie dans les langues, avant d'examiner ses implications dans la confection de dictionnaires au Gabon. Pour mener à bien ce travail, nous nous référons exceptionnellement au dialecte fang ntoumou de Bitam, pour ce qui est naturellement applicable aux autres langues du Gabon.

Mots clés: DICTIONNAIRES, EXPERTS, LANGUES, LEXICOGRAPHIE, LOCUTEURS, MAIN INVISIBLE

Abstract: The "Invisible Hand" in Languages and the Compilation of Local Dictionaries. Although it is not their intention, it seems that dictionary users are often guided by *an invisible hand* similar to the one developed in economy by Adam Smith. By seeking individual performance, they end up contributing to the development of their languages without being their intention. To Adam Smith, it is by seeking personnel development that people develop the nation as well. According to him, it is *an invisible hand* that leads them to reach that purpose which was not their intention, and as a result they turn out to be more profitable to society than if it had been their intention to make that sort of contribution. Regarding these similarities between the manifestations of *the invisible hand* in economy and in languages, this paper is an attempt to enunciate the same theory in languages. Using a theory from the field of economy to explain facts in languages is not a new approach. The French sociolinguist Louis-Jean Calvet founded a similarity between the law market which determines the exchange value of a currency, and the respective values of the differ-

ent languages in the world which correspond to their degree of diffusion. Referring to the invisible hand in economy, we discuss first the enunciation of the same theory in languages, before analyzing the implications for the compilation of dictionaries in Gabon. This paper refers specifically to Bitamfang ntoumou dialect, for an analysis which is obviously also relevant to other Gabonese languages.

Keywords: DICTIONARIES, EXPERTS, INVISIBLE HAND, LANGUAGES, LEXICOGRAPHY, SPEAKERS

Introduction

L'application d'une théorie économique afin de démontrer certains faits de langues n'est pas une nouveauté. Le sociolinguiste français Louis-Jean Calvet (2004), a en effet déjà usé de la même démarche. Il s'est appuyé sur la loi économique du marché, dans lequel la force d'une monnaie dépend de sa valeur d'échange, pour essayer d'expliquer la différence entre les valeurs des différentes langues du monde en fonction de leur degré de diffusion. Discuter des mêmes causes et effets observés dans les langues et similaires à *la main invisible* en économie, nous taraudait l'esprit. La démarche de Calvet nous a donné de l'assurance.

Pour l'économiste et philosophe écossais du siècle des Lumières Adam Smith (1723–1790), les activités économiques ne sont pas fondamentalement générées par l'intervention de l'Etat. Pour lui, ces activités proviennent plus en grande partie du produit de l'ensemble des actions qui emmènent les individus qui sont comme mus par une sorte de *main invisible*, vers la recherche de l'intérêt et du profit personnel, que celles provenant des activités menées par l'Etat. Combinées aux mécanismes du marché, ces activités menées par les individus permettent à chacun d'eux, de contribuer non seulement à s'enrichir personnellement, mais également à accroître la richesse de l'Etat. Pour cet économiste, la recherche de l'intérêt individuel est alors le plus sûr moyen d'œuvrer pour accroître la richesse des nations.

Pourquoi pensons-nous qu'il est possible d'évoquer *une main invisible* dans les langues semblable à celle développée par Adam Smith en économie? Nous partons du constat que la langue n'est pas fondamentalement générée par l'intervention des experts. Nous ferons ici particulièrement allusion aux linguistes et lexicographes pour ne citer que ceux-là. Ces derniers, ne font que signaler, représenter, standardiser et analyser les activités linguistiques des locuteurs et les mécanismes linguistiques qui les sous-tendent. Le fonctionnement et le développement d'une langue, proviennent donc de l'ensemble des activités des locuteurs résultant des milliers de décisions spontanées individuelles que ceux-ci prennent dans leurs incessants besoins de communiquer. Combinées aux mécanismes linguistiques, ces activités permettent à chacun des locuteurs, de contribuer non seulement à mieux performer personnellement, mais également à enrichir la langue de leur communauté, sans que ce cela soit leur intention. La recherche de la performance individuelle, peut alors être perçue comme le plus

sûr moyen d'œuvrer pour développer les langues. Ne voyons-nous pas dans ces faits empiriques comme les manifestations *d'une main invisible* dans les langues semblable à celle énoncée par Adam Smith en économie? Ce sont ces faits empiriques qui nous ont fait germer l'idée d'une énonciation dans les langues d'une théorie de *la main invisible*, semblable à celle développée en économie par Adam Smith. Cette théorie aura donc nécessairement des implications en lexicographie. Nous allons essayer de les analyser en nous référant à la confection de dictionnaires au Gabon.

Pour faire ressortir l'intérêt, les objectifs et les enjeux de ce travail, nous commencerons d'abord par présenter brièvement la théorie de *la main invisible* telle qu'énoncée par Adam Smith. En nous y renvoyant, nous discuterons ensuite de la proposition d'énonciation de la même théorie dans les langues. Nous terminerons ce travail par une analyse des implications de cette théorie en lexicographie à partir de certaines données de la confection des dictionnaires bilingues monodirectionnels que nous conduisons au Gabon. Pour mener cette analyse, nous nous référons exceptionnellement au dialecte fang ntoumou de Bitam, pour ce qui est naturellement applicable aux autres langues du Gabon.

1. "La main invisible" en économie

1.1 La théorie

La main invisible est une expression célèbre que nous devons à Adam Smith, économiste anglais, considéré comme le père de l'école libérale. Il a décrit comment par la recherche de l'intérêt individuel et du profit et par les mécanismes du marché, chaque citoyen contribue non seulement à s'enrichir personnellement mais également à accroître la richesse de la collectivité. Pour lui, la recherche de l'intérêt individuel est le plus sûr moyen d'œuvrer pour accroître la richesse des nations et le rôle de l'État doit être strictement limité.

C'est de la même façon que nous pensons que la recherche de la performance linguistique individuelle est le plus sûr moyen pour accroître la richesse des langues. De ce fait, le rôle des experts, en particulier des linguistes fondamentalistes et descripteurs, qui est tout aussi important, ne doit être toutefois limité qu'à signaler et à standardiser cette performance. Voici la théorie de *la main invisible* telle qu'énoncée par Adam Smith:

Ce n'est que dans la vue d'un profit qu'un homme emploie son capital. Il tâchera toujours d'employer son capital dans le genre d'activité dont le produit lui permettra d'espérer gagner le plus d'argent. (...) A la vérité, son intention en général n'est pas en cela de servir l'intérêt public, et il ne sait même pas jusqu'à quel point il peut être utile à la société. En préférant le succès de l'industrie nationale à celui de l'industrie étrangère, il ne pense qu'à se donner personnellement une plus grande sûreté; et en dirigeant cette industrie de manière que son produit ait le plus de valeur possible, il ne pense qu'à son propre gain; en cela, il est conduit par une main invisible, à remplir une fin qui n'entre nullement dans ses intentions; et ce n'est pas toujours ce qu'il y a de plus mal pour la société, que cette fin

n'entre pour rien dans ses intentions. Tout en ne cherchant que son intérêt personnel, il travaille souvent d'une manière bien plus efficace pour l'intérêt de la société, que s'il avait réellement pour but d'y travailler.

1.2 Implications de la théorie

La théorie énoncée par Adam Smith a des implications dans le fonctionnement de l'économie. Pour lui, c'est tout d'abord une façon élégante de souligner que l'État n'a pas à se mêler de la vie économique et que les mécanismes du marché valent bien toutes les lois du monde. Le fonctionnement du marché trouve à sa base des milliers de décisions individuelles où chaque acteur économique cherche naturellement les moyens de s'enrichir personnellement. Ensuite, l'expression d'Adam Smith est bien plus qu'une simple métaphore: elle résume un programme idéologique, celui qui vise à faire du marché l'unique régulateur de l'ensemble de la vie économique. C'est pourquoi l'expression est si fréquemment reprise, par ceux qui s'opposent à ce programme, aussi bien que par ceux qui en sont partisans.

2. Les raisons d'"une main invisible" dans les langues

Comme nous l'avons déjà dit dans nos propos liminaires, il y a des faits empiriques qui nous laissent croire qu'il pourrait bien exister *une main invisible* dans les langues similaire à celle qui est en économie. Nous constatons que le locuteur est poussé vers la recherche de la performance individuelle dans le souci de mieux communiquer et, crée involontairement par cette voie, des formes de la langue qui sont ensuite attestées par la communauté entière. En fang ntoumou de Bitam, ce sont les locuteurs qui ce dynamisme linguistique, ont offert aujourd'hui à leur langue des mots tels que *nsômono* (littéralement *argent* nu pour signifier *argent liquide ou espèces*), *Sikolo* (école), *Sitogho* (chaussette), *angong* (cadenas), *ngaa* (fusil), *biyéne* (lunettes), *messisse* (allumettes), *Dokira* (Docteur), *nlohaoussa* (littéralement *tête d'Haoussa* pour signifier *bouilloire*, à cause de la ressemblance entre le couvercle de cet ustensile et la coiffure des peuls du Cameroun), etc. Ce sont là des mots auxquels les locuteurs n'auraient certainement pas pensé.

Comme en économie, cette recherche de la performance individuelle n'est pas toujours ce qu'il y a de plus mal pour les langues, que cette fin n'entre pour rien dans les intentions de ce dernier. En effet, on constate que tout en ne cherchant que son besoin personnel de communiquer, l'individu contribue souvent d'une manière bien plus naturelle, donc plus efficace pour l'intérêt de la langue, que s'il avait réellement pour but d'y contribuer.

Toutefois, Il est très important qu'on nous comprenne à notre juste propos: nous ne voulons aucunement exclure les experts des questions de langues de la vie de celles-ci en général, et de celles locales du Gabon en particulier. Nous voulons simplement signifier comme en économie, que l'intervention de l'Etat, qui pourrait être assimilée à celle des experts dans les langues, est bien évidem-

ment nécessaire mais seulement à certains égards. Les experts linguistes font ressortir et analysent les mécanismes linguistiques que les locuteurs utilisent inconsciemment, tandis que d'autres experts élaborent des ouvrages de lecture, des grammaires et des dictionnaires par exemple. Ils ne se substituent pas toutefois aux locuteurs. Ils ne font que décrire l'usage que ces derniers font de leur langue dans l'accomplissement de ses fonctions de communication. Cheik Anta Diop (1979: 345), dit que la langue est une entreprise de longue haleine où les modalités des locuteurs tout comme celle des experts ne s'excluent pas du tout.

Il est donc question dans le traitement des langues locales du Gabon dont il est question dans ce travail, que les experts, comme dans les langues qui essaient avec bonheur aujourd'hui telles que l'anglais et le français par exemple, respectent le fait que les manifestations orales et écrites de ces langues ont déjà été fécondées par les locuteurs. Ils ne leur restent plus qu'à les signaler, les représenter, les étudier et à les standardiser afin de formaliser et de démocratiser leur usage.

2.1 Les locuteurs et l'écriture française dans l'usage local

Le besoin de communiquer et la recherche des moyens pour y parvenir sont naturels chez l'homme. Ce n'est qu'un aspect d'une caractéristique qui lui est générale: la recherche de l'intérêt personnel et la nécessité vitale de s'adapter à son environnement. Dans le souci de représenter les manifestations écrites locales pour les besoins du développement du Gabon¹, les administrateurs locaux et français se sont naturellement appuyés sur l'orthographe et l'alphabet latin de l'écriture française, qui a certes intégré la société gabonaise par la force de la colonisation. Nous considérons l'écriture comme "la liste de signes graphiques pouvant être employés pour la notation de manifestations écrites", cf. Chiss et Puech (1983: 18). Ce sont donc les milliers de décisions individuelles prises par les locuteurs depuis la période coloniale dans le souci personnel... *de domestiquer la science et la technologie, de défendre et de faire rayonner la culture, en un mot aux besoins suscités par des conceptions ou des pratiques qui apparaissent*²... qui ont fourni les manifestations écrites locales existant aujourd'hui.



Figure 1: 1960, le visage du Gabon indépendant avec le 1er Président gabonais Léon Mba et son Vice-président Albert Bernard Bongo. Photo: Internet

Le Gabon indépendant en 1960, se devait de se moderniser, et s'appuyer sur l'écriture française était l'approche la plus plausible qui soit pour se doter d'une graphie locale moderne. L'écriture d'une langue ne sort nullement toute armée des cerveaux de quelques démiurges. Elle émerge des locuteurs eux-mêmes qui s'approprient des signes usuels qu'ils soient locaux ou allogènes provenant de leur environnement immédiat. Toutefois, il est vrai que les locuteurs doivent attendre l'autorité extérieure des experts pour déclarer le caractère officiel de l'usage des signes d'une écriture. Pendant un temps, les premiers ont en effet l'intuition que la graphie qu'ils utilisent n'est pas légitime, si l'autorité extérieure des experts n'en a pas encore déclaré le caractère officiel. C'est le cas en fang ntoumou de Bitam avec les noms des différents villages et tribus tels qu'*Esseng*, *Essabeigne*, *Effack*, *Nkolmengoua*, *Biléossi*, *Medoumou*, etc. Ces noms qui sont au départ souvent griffonnés par les locuteurs sans beaucoup d'assurance, sont repris tels quels par les officiers de l'administration civile et d'Etat. Les locuteurs sont souvent surpris de constater que leur graphie est prise en compte.

Le processus de la standardisation d'une écriture est l'inverse de qui est communément admis. Les experts l'extraient des formes courantes de la graphie germée par les locuteurs. Ce n'est donc pas en réalité une minorité compétente d'experts qui statuent en tant que tel sur l'écriture d'une langue. Ces derniers ne font que standardiser ce qui remonte à la surface des milliers de décisions individuelles prises par les locuteurs face aux besoins de représenter leur langue. Cela atteste les propos de Cheikh Anta Diop (1979: 345) déjà cités en supra en rapport avec les implications respectives des locuteurs et des experts dans l'entreprise langue. Au Gabon, les locuteurs ont donc comme mus par *une main invisible*, instrumentalisé l'écriture française présente dans leur environnement immédiat, afin de transcrire les langues locales du Gabon pour diverses raisons prosaïques. Cela a donné naissance à une graphie que nous essayons de présenter dans ce corpus ci-dessous. Se voulant être représentatif, ce corpus montre les transcriptions de certains mots courants désignant des essences locales, des groupes ethnolinguistiques, des noms de villes et de personnes:

Padouk, Ekouk, Bifoun, Ngozè, Mouvingui, Sorro, Moabi, Douka, Doussié, Doussala, Matsanga, Ndzigou, Obiang, Mvet Ou Mvett, Moussoungou, Fougamou, Malamba, Mboudi, Bwiti, Ndjémbè, Gnèmbwè, Mpassa, Ngouoni, Révignet, Azizetngounié, Mvoung, Ovang, Megwang, Nyanga, Nkolmengoua, Mayumba, Oyogouyogou, Melhogouane, Dufule, Mbamisoghè, Nkolmessass, Mpouna, Adoua, N'lodè, Bouénguidi, Doya, Megwa, Okwè, Woleu, Ntem, Assala, Ogouée, Kougueleu, Rompavet, Ndjolé, Essassa, Mindoubé, Bindoumessang, Woubélé, Doussala, Koulamoutou, Minkwe, Boukandou, Akéndengué, Ekang, Eyang, Bakoumba, Ndong, Mavoungou, Ovan, Nzeng, Oyoubi, Nzouba, Bibang, Moundounga, Ndinga, Mpouhot, Nzengue, Nzengui, Medouneu, Mindoungani, Tchibanga, Biboulou, Mboudi, Okoumé, Ntoutoume, Iboundji, Mitzic, Moussavou, Assoumou, Padouk, Kevazingo, Bounguendza, Mimongo, Massango, etc.

*Le dictionnaire des noms propres du Gabon*³ ainsi que *Les plantes utiles du Gabon: l'essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*⁴ parmi tant d'autres ouvrages de référence, confirment comment locuteurs eux-mêmes sont à l'origine de l'écriture de leurs langues pour diverses raisons concrètes. Il ne reste plus aux experts qu'à signaler les formes graphiques qui sont fréquentes et en extraire l'écriture des différentes langues locales du Gabon.

2.2 Les locuteurs et la terminologie française dans l'usage local

Ce qui vient d'être démontré dans le cas de la graphie des langues locales du Gabon, insinuée à la base par les locuteurs eux-mêmes, est exactement ce qui se produit également dans la terminologie. Il n'est donc pas nécessaire de reprendre une autre démonstration qui ne se distinguera qu'à quelques variables près avec celle qui a déjà été réalisée plus haut. Nous pouvons seulement dire que la terminologie locale moderne actuelle, émane de milliers de décisions individuelles d'emprunter en grande partie au français⁵, les mots manquant dans les langues locales dans le souci des locuteurs locaux d'accroître leur seule performance personnelle. Dans la nécessité de s'adapter à la société moderne qui est sans cesse en mutation, les locuteurs doivent intégrer des dizaines de milliers de mots et les concepts que ces derniers véhiculent. Ces mots sont liés à l'économie, à la politique, au social, à l'administration, à la justice, aux finances, au commerce, à l'industrie, aux sciences, aux techniques, à l'informatique, à l'électronique, à la mécanique, au sport, à la drogue, à la violence, à l'armement, à la culture, à l'histoire, à l'environnement, l'art, la musique, l'architecture, etc. Certains mots ont été intégrés tels qu'ils ont été empruntés, quelques uns ont été modifiés pour se conformer aux structures linguistiques locales et d'autres ont été remplacés par des termes locaux.

C'est donc comme mus par *une main invisible* que les locuteurs ont eux-mêmes initié l'enrichissement et la modernisation de leurs langues locales, car il faut toujours rappeler que cela ne rentrait pas dans leurs intentions au départ. La terminologie d'une langue ne naît pas des créations lexicales provenant des experts. Comme avec l'écriture, les locuteurs doivent seulement comme attendre l'autorité extérieure des experts pour déclarer le caractère officiel de l'usage d'un nouveau mot local, un mot étranger ou qui en résulte. En fang ntoumou de Bitam, ce sont les cas par exemple de *mema* (maman), *pepa* (papa), *nsengavion* (aéroport), *nsengndama* (stade), *ndama* (ballon, football), *banque*, *ordinateur*, *université*, *ministre*, etc. Les locuteurs ont en effet pendant un temps, l'intuition que l'usage de ces mots n'est pas légitime, si l'autorité extérieure des experts n'en a pas encore déclaré le caractère officiel. Comme avec l'écriture, ils sont souvent donc assez surpris de voir que c'est leur usage qui est décliné par les experts. Ils penseraient qu'ils auraient mal intégré *maman* (*mema*), *papa* (*pepa*), ou qu'ils n'auraient même pas dû le faire au départ, ou qu'ils ne devraient pas utiliser les mots français *université*, *ministre*, etc., qui sont allogènes au fang

ntoumou de Bitam.

Le processus de la standardisation des nouveaux mots est semblable à celui de l'écriture. Il est l'inverse de qui est communément admis. Qu'un nouveau mot, local ou étranger, entre dans une langue, ne signifie pas que cela est l'œuvre d'une minorité compétente d'experts. Ces derniers ne font que le signaler et en démocratiser l'usage.

3. La théorie de "la main invisible" dans les langues

En remplaçant les variables économiques par les variables linguistiques dans la théorie de *la main invisible* d'Adam Smith, on obtient une énonciation de *la main invisible* dans les langues. Comme dans les activités économiques, nous pensons que c'est dans la recherche de la performance individuelle à travers les mécanismes de la langue, dont les locuteurs n'ont pas conscience, que chacun d'eux contribue non seulement à s'enrichir personnellement, mais également à développer la langue qui va servir à la collectivité toute entière. Ils utilisent inconsciemment l'*apocope* par exemple, pour faire jouer à la langue sa fonction poétique, une des six fonctions du langage développées par Jakobson (1963). C'est ainsi qu'en fang ntoumou de Bitam, dans *éfassé tsite* (gigot de viande) et *éfousse si* (lopin de terre), *éfassé* (gigot) et *éfousse* (lopin) perdent leurs phonèmes /sse/, et on dit *éfa tsite* et *éfou si* pour une meilleure sonorité ou poésie de la langue. Ils peuvent encore utiliser ce qui est appelé l'*emprunt direct* analysé par des auteurs tels que Madiba (2001: 64-74), en prenant directement dans le français en particulier, les mots dont ils ont besoin dans leur langue. Il est donc naturel d'entendre les locuteurs du fang ntoumou de Bitam, intégrer les mots tels que *Ministre, Député, Sénateur, Ciment, Brique, Carreau, Rideau, Glaçon, Glace, Université, Lycée, Banque, Crédit, Télévision, Dictionnaire, Ordinateur, Téléphone*, etc. dans leur conversation: *E mwane wôme a ne Ministre* (Mon enfant est Ministre), *Ma fwiri be carreaux e nda dzame* (Je place les carreaux dans ma maison), *E bwane bâme be ne (ou) ba ke o Lycée ya Université* (Mes enfants sont/vont/fréquentent le Lycée et l'Université), etc.

Ces faits de langue qui témoignent de l'existence d'une main invisible dans les langues, démontrent que la recherche de la performance individuelle est le plus sûr moyen d'œuvrer pour développer les langues des nations. En cela, le rôle des experts doit être strictement limité qu'à la description des langues telles qu'elles se pratiquent par les locuteurs. La théorie de la main invisible dans les langues s'énoncerait alors comme suit:

Ce n'est que dans la vue de la communication qu'un homme emploie ses habilités à performer. Il tâchera toujours d'employer son habileté et sa créativité dans un aspect de la communication dont la performance lui permettra de mieux communiquer. (...) A la vérité, son intention en général n'est pas en cela de servir l'intérêt public, et il ne sait même pas jusqu'à quel point il peut être utile à la société. En se focalisant sur le développement de sa langue à celui de la langue étrangère, il ne pense qu'à se donner personnellement une plus grande capacité

de communiquer; et en dirigeant cette performance de manière que sa capacité à communiquer ait le plus le plus de valeur possible, il ne pense qu'à sa propre communication; en cela, il est conduit par une main invisible, à remplir une fin qui n'entre nullement dans ses intentions; et ce n'est pas toujours ce qu'il y a de plus mal pour la langue, que cette fin n'entre pour rien dans ses intentions. Tout en ne cherchant que son intérêt personnel, il travaille souvent d'une manière bien plus efficace pour l'intérêt de la langue, que s'il avait réellement pour but d'y travailler.

3.1 Implications de la théorie

L'énonciation de la théorie de *la main invisible* dans les langues, a les mêmes incidences que celle énoncée en économie. C'est une façon élégante de souligner que les experts n'ont pas à trop se mêler de la vie des langues, et que les mécanismes et phénomènes naturels de la langue, dont les locuteurs sont à la source, valent bien toutes les théories sur les langues du monde. Dans le traitement des langues locales du Gabon du Gabon, on remarque que les experts, en l'occurrence les linguistes africanistes gabonais, les ont pris en otage. La graphie germée par les locuteurs depuis la période coloniale est jugée non-scientifique et basée sur l'écriture du français qui ne transcrirait pas ces langues de façon exacte et cohérente. Par ailleurs, les mots français y sont considérés comme une impureté alors que ces derniers permettent à ces langues d'accomplir pleinement et sans ambiguïté leurs différentes fonctions de communication. L'histoire a voulu que le français soit aujourd'hui la langue officielle du Gabon, donc une des langues de ce pays, et point d'ancrage des pratiques idéologiques et culturelles. On ne peut réécrire l'histoire. On l'assume. La question de l'usage inconscient et naturel par les locuteurs français dans les langues locales du Gabon ne devrait plus faire l'objet de débats. Surtout, lorsqu'on sait que cette langue s'étant ouverte au monde, véhicule une culture universelle. En instrumentalisant naturellement le français dans ces langues, les locuteurs les rendent fonctionnelles et ce, toute en harmonie et en adéquation avec l'environnement immédiat. Ils les ouvrent aussi au monde et les évitent d'être repliées sur elles-mêmes. Cette caractéristique d'être ouvertes au monde via le français les protégera de la stagnation, de l'étiollement et de la disparition.

Le fonctionnement de la langue, aussi bien à l'oral qu'à l'écrit, trouve à sa base, des milliers de décisions individuelles où chaque locuteur cherche naturellement les moyens personnels de communiquer. C'est ce dynamisme dont les locuteurs sont l'origine, et auquel s'ajoute bien évidemment l'intervention des experts, qui à la base fournit à la langue ses propriétés fondamentales de fonctionnement et de vie. L'expression de *main invisible* que nous venons de développer dans les langues, sera comme celle d'Adam Smith en économie. C'est-à-dire, qu'elle sera aussi bien plus qu'une simple métaphore: elle va résumer un programme idéologique, celui qui vise à faire de la performance individuelle des locuteurs et les milliers de décisions qu'ils prennent comme l'unique générateur de l'ensemble des manifestations orales et écrites de la langue. C'est

pourquoi nous pouvons dire que l'expression qui est développée ici sera reprise comme *la main invisible* en économie, par ceux qui s'opposeront à ce programme, aussi bien que par ceux qui en seront partisans.

3.2 Les locuteurs et les experts comme modalités complémentaires

La promotion et la valorisation des langues locales du Gabon sont aujourd'hui plus une question d'urgence que d'actualité tant elles pèsent sous une forte menace de disparition. Dans le Projet de Plan Décennal de l'Enseignement, de la Formation et de la Recherche, l'introduction des langues locales comme activités d'immersion a été retenue comme actions et mesures⁶. De nombreux travaux linguistiques et lexicographiques avec l'objectif de stimuler et soutenir l'intégration de ces langues dans les programmes d'enseignement ont déjà été élaborés ou sont en cours de projet. La fondation Raponda-Walker a édité des ouvrages d'apprentissage des langues locales du Gabon⁷ qui, bon en mal en, sont utilisés dans les établissements qui intègrent ces langues comme enseignements. Nous suggérons que tout comme en économie, où la création et la production de l'individu jouent un rôle prépondérant, que le traitement des langues locales décrive l'usage courant qu'en font les locuteurs. C'est cet usage qui est véritablement indicatif de la réalité de la situation linguistique du Gabon, et également de la façon dont ces langues remplissent leurs fonctions de communication. Ce sont ces données que les experts doivent faire ressortir, qu'elles leur plaisent ou non.

Les locuteurs qui comptent parmi eux les dépositaires de la mémoire collective, autrement dit, les administrateurs, les scientifiques, les hommes de pensée, les artistes chanteurs et compositeurs, les journalistes, les éducateurs, les missionnaires et les responsables politiques de chaque groupe ethnolinguistique, etc. ne sont pas à ignorer et à exclure dans la vie des langues locales. Tout au contraire. Ce sont ces personnes là qui, contre toute attente, rêvent et façonnent, par quelque souci de fantasme culturel, un trésor de la culture, non pas française, mais gabonaise (nous nous référons à Kalonji 1993: 79). Des intellectuels et des génies dans leur genre à base culturelle locale, elles se sont harmonieusement intégrées dans les normes linguistico-culturelles modernes par le biais de la culture française. Ces personnes là s'avèrent dans la plupart des cas, les aptes défenseurs des langues locales du Gabon dans le nouvel ordre sociopolitique et culturel (nous nous référons encore à Kalonji *idem*). Dotés de la double identité franco-gabonaise, les dépositaires de la mémoire collective, sont ceux qui mieux que quiconque, assument et peuvent assumer harmonieusement l'héritage des langues locales en rapport avec les exigences de la société moderne. Ils intègrent quelques mots tels qu'empruntés au français ou d'autres langues étrangères, modifient certains par rapport à la structure linguistique de leurs langues locales respectives et remplacent les mots empruntés par les mots locaux lorsqu'ils se sont familiarisés avec le concept ou l'objet véhiculé par l'emprunt. Les experts doivent décrire cette façon dont les langues sont ense-

mencées par les locuteurs et non de prescrire ce qui selon eux, serait la véritable terminologie des langues locales. Comme l'a souligné Mojela (2014), dans les langues africaines, il semble que les experts veuillent standardiser ce que les locuteurs ne parlent pas, et ils veulent faire parler ce que les locuteurs n'ont pas standardisé.

Ces locuteurs ont aussi la capacité de générer une écriture à partir d'une graphie s'appuyant sur celle du français pour représenter les manifestations écrites locales. C'est cette graphie qui jette les bases du développement d'une écriture locale qui harmonieusement assure le relais de l'oral, la préservation et la diffusion des langues locales du Gabon. Nous ne parlons pas ici d'une transcription phonétique, algébrique, exacte provenant de collectes et d'analyses de type universitaire et qui rend cette écriture locale scientifique. Nous parlons d'une écriture locale, qui résulte du dynamisme des intellectuels franco-gabonais qui transcrivent les langues locales pour des raisons prosaïques. C'est cette graphie qui tient compte des aspirations, des besoins des locuteurs, et qui est portée par les apports aussi bien locaux qu'étrangers en corrélation avec le contexte sociopolitique et culturelle qui leur est contemporain. Si la production des formes écrites générée par les locuteurs n'est pas faite dans l'objectif de la recherche d'un idéal linguistique, elles ont toujours toutefois constitué des données cohérentes et fonctionnelles qui servent de support d'analyses et d'études pour les experts. Au Gabon, ce sera donc une écriture produite à partir de la graphie fécondée par les locuteurs *qui offrira à la linguistique africaniste locale les voies d'accès à la compréhension du devenir des langues locales, de leur constitution, de leur diffusion culturelle, à une prise sur les questions de l'histoire, de la littérature, de la politique, des différences sociales.*

Certes, la graphie produite par les locuteurs ne représente pas tous les sous-systèmes phonologiques des langues locales. Elle ne défigure et ne travestit pas également celles-ci. Comme dans toutes les langues qui essaient avec bonheur aujourd'hui, c'est bien à partir de la graphie des locuteurs que *les propriétés fondamentales de la langue* ont toujours été *rendu visibles et tangibles* (Nous utilisons les propos de Chiss et Puech (1983: 8). Les locuteurs vont apparaître la graphie de leur langue pour des raisons prosaïques de comptabilité, de commerce, etc. indépendamment de toute recherche de connaissance linguistique. Celle-ci comporte donc parfois des bizarreries par rapport à une transcription rigoureuse. Toutefois, il ne revient pas aux experts de les ajuster à une quelconque cohérence et exactitude scientifique ou théorique. Cela n'a jamais été le cas dans les langues qui sont considérées comme les plus dynamiques. Les experts ne se limitent qu'à les signaler et à les représenter du moment où elles font partie de l'usage courant des locuteurs.

En français par exemple, il y a des lettres nulles dans les mots tels que **corps** et **doigt**; **e** se prononce **a** dans **femme** et **solennel**; **eu** se dit **u** dans **j'ai eu**, **tu as eu**, etc.; **c** se lit **g** dans **second**; **ch** s'énonce **k** dans **eucharistique**, **psychologie**, **écho**, **chrétien**, **chœur**; **chrome**, **chorale**, **chorégraphie**, **chlore**, **x** est lu **s** dans **six**, **dix**; etc. Ces bizarreries sont aussi observables dans les mots en

langues locales du Gabon. *Mitzic*, Chef lieu du Département de l'Okano dans la province du Woleu-Ntem se lit *Mindzik* [mĩdzík], *Makokou* se dit *Makòkou* [makòkú] dans la langue Ikota et *Mekoughe* [mæk^húgə] dans le dialecte fang nzaman. *Ondo*, *Ndong* et *Nguéma* noms propres de personne en fang se disent respectivement *Ondòò* [óndò:], *Ndong* [ndõ:] et *Nguéma* [ngema] dans le dialecte fang ntoumou mais se lisent *Oundò* [úndò], *Ndoueng* [ndüə] et *Ndjéma* [ndzómà] en fang nzaman. *Ella*, nom propre de personne en fang, se lit *Éllaa* [élà:] dans le dialecte fang ntoumou de Bitam, *Éllèè* [élè:] dans le fang ntoumou d'Oyem et *Illa* [ílà] dans le fang nzaman. En fang nzaman, *Ékiéma*, *Nguere*, *Benga* et *Epégnè* se lisent *Itchiéma* [ítjémà], *Ndjere* [ndzòRè], *Behang* [bóã] et *Ékpain* [ékpè].

Les experts ne produisent pas la langue que ce soit dans sa forme orale ou écrite. Ils ne font récolter et répandre ce que produisent les locuteurs, tels les apiculteurs récoltent et répandent le miel par la suite. Par ailleurs, une langue est la mémoire et le vécu d'un peuple. Les locuteurs gravent naturellement et inconsciemment dans leurs langues tous les aspects matériels ou abstraits de leur environnement et le cours de leur histoire, et garantissent à ces dernières sa fonction de repère et de mémorisation de leur peuple. Une intervention non mesurée des experts, va dénaturer et déstabiliser ce processus inconscient de la vie d'une langue dans lequel les locuteurs participent activement et qui fait que la langue soit la mémoire et le vécu de son peuple. La vie d'une langue peut être représentée sous la forme d'une chaîne linguistique, qui est un cycle dans lequel les locuteurs constituent le point de départ en produisant la langue et les experts la standardisant ensuite.

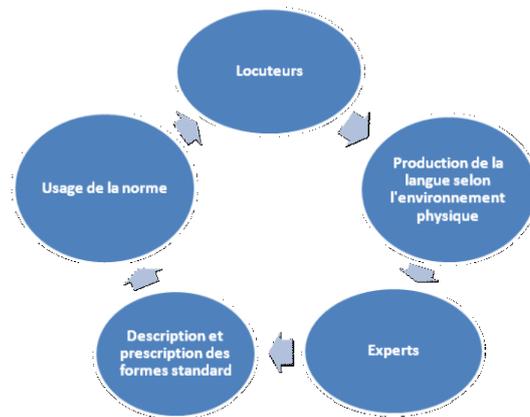


Figure 2: La chaîne linguistique

4. La main invisible et la confection de dictionnaires locaux

4.1 Les locuteurs et l'acquis de l'écriture française dans l'usage local

Fort de tout ce qui a été précédemment dit, la graphie insinuée par les locu-

teurs constitue un acquis dans l'élaboration de l'écriture des différentes locales du Gabon. Cette graphie est très influencée par le français par l'utilisation de son orthographe et celle de l'alphabet latin, car la langue de Molière fait partie du patrimoine linguistico-culturel du Gabon. C'est un processus inconscient et un facteur constitutif rentrant dans le fonctionnement et le développement d'une langue quand elle côtoie une autre. Les locuteurs ont donc déjà jeté les bases de l'écriture locale qui est influencée par celle du français. Cela ne constitue aucun mal. Le signe linguistique est arbitraire et l'usage d'une écriture allo-gène n'a jamais altéré l'identité d'une langue et de sa culture. Cette graphie est déjà inscrite dans le complexe psychologique des populations locales et établie dans les différents groupes linguistiques du Gabon. C'est aussi celle qui est utilisée dans l'écriture officielle des actes administratifs et gouvernementaux. L'écriture qui y découlera sera donc porteuse de stabilité, d'harmonisation et de fonctionnalité immédiate. Et comme le dit Saussure (1964: 98-101), *il n'est pas au pouvoir de l'individu de rien changer à un signe une fois établi dans un groupe linguistique*. Vouloir créer une autre écriture qui serait propre aux langues du Gabon pourra donc être plus porteur de déstabilisation qui conduira à des impasses qu'un aboutissement.

Dans un dictionnaire bilingue monodirectionnel français et fang ntoumou de Bitam que nous simulons ici, le guide de l'utilisateur présentera donc un alphabet tiré de la graphie des noms propres et noms des localités administratives transcrites dans ce dialecte. La façon dont les mots sont prononcés en fang ntoumou de Bitam à partir de l'écriture de cette langue qui, comme celles des autres langues du Gabon, est influencée par l'orthographe et l'alphabet latin du français, sera indiquée. Les données de ce dictionnaire simulé sont extraites du dictionnaire du même type que nous confectonnons pour la langue fang. Elles ne sont pas encore abouties. Elles sont donc encore perfectibles.

L'ALPHABET FANG NTOUMOU DE BITAM (une proposition)

Il se composerait de 26 lettres:

A	a	H	h	O	o	V	v
B	b	I	i	P	p	W	w
C	k	J	j	Q	q	X	x
D	d	K	k	R	r	Y	y
E	e	L	l	S	s	Z	z
F	f	M	m	T	t		
G	g	N	n	U	u		

Et les lettres particulières au fang seraient: **gw, mv, mf, mv, nd, ng, nk** et **nl**.

LA PRONONCIATION

De façon générale, la majorité des voyelles finales sont longues en fang: *Songo*

(jeu traditionnel, lire *son-hôô*), *Ondo* (nom propre, lire *Ondôô*), *Edou* (nom propre, lire *Edouou*), *Eyang* (nom propre, lire *Éyaang*), *Élla* (nom propre, lire *Éllaa*), *Éko* (nom propre, lire *Ekôô*), *Ndong* (nom propre, lire *Ndoong*), *bo* (cerveau, cervelle, lire *bôô*), *ébo* (pourri, lire *ébôô*), *lo* (serrure, lire *lôô*), *mongo* (enfant, lire *mon-hôô*), *a gni* (rentrer, lire *a gnii*), *a so* (douter, lire *a soo*), *dza* (village, lire *dzaa*), *a wui* (désherber, lire *a wuii*), *aba* (corps de garde, lire *abaa*), *kou* (tortue, lire *kouou*), *nkô* (corde, lire *nkôô*), *nko* (colline, montée, montagne lire *nkoo*), *a sông* (provoquer, lire *a son-hôô*), *kà* (sœur, lire *kàà*), *mbi* (course, lire *mbii*), *mbî* (terrier, lire *mbîi*), *abé* (kola, lire *abéé*), *abè* (accouchement, lire *abèè*), *essi* (cheveu, lire *essi*), *ze* (barbe, lire *zee*), etc.;

Voyelles: a, e, i, o, ui

- [a] est noté *a*: *abàme tagha* (paquet de cigarette), *abàme* (fil à coudre/tresser), *alou* (jour, nuit), *abègne* (cuisse), *abègne* (sein), *dzi* (repas), *amang* (joue), *abô* (pied), *aboume* (ventre), *alô* (oreille), *agnou* (bouche), *avwate* (filet), *akane* (fesse), etc.;
- [ɔ̃] est noté *ong*: *Ndong*, *ndzong* (route), *wong* (peur), *wông* (anguille), *nlông* (liane), *nlong* (ligne), *kong* (malédiction, sort), *zong* (aubergine), *abong* (genou), *mbông* (manioc), *avong* (graisse), etc.;
- [ã] est noté *ang* et généralement en fin de mot: *Obiang*, *ébang* (paquet de bois, manioc, argent), *a bang* (aiguiser), *a kang* (griller ou faire sauter en cuisine), *a dzang* (disparaître), *mang* (mer), *akang* (fumoir), *amang* (joue), *biang* (médicament), etc.;
- [ɔ̃] est noté *ô* pour l'opposer *o* dans des homographes: *fope* (bavard, e), *nlô* (mouche) ≠ *nlo* (tête), *ntôke* (mince) ≠ *ntoke* (cuillère), *nkô* (corde) ≠ *nko* (colline, montagne), etc.;
- et il est parfois noté *ö* pour séparer des homophones, *ntôke* (mince) ≠ *ntoke* (cuillère) ≠ *ntôke* (tresse), etc.;
- [ɛ] est aussi noté *è*: *zègne* (faim), *gniè* (mère), *essègne* (travail), *abègne* (sein), *à wuègne* (rire, sourire), *luègne* (rire, sourire), *kuègne* (singe), *abègne* (cuisse), *édzakasse* (chameau), *ékamelôk* (cheval), etc.;
- [e] est noté *é*: *élé* (arbre), *édzina* (chambre), *édoke* (profond, e), *éfousse* (morceau de viande, pain, terre), *éfassé* (morceau de verre et bois), *ébasse* (pâte), *ébasse owono* (pâte d'arachide), *édoudouk* (sueur), *ébirane* (désordre, désordonné, e), *ékiègne* (fer), *éloume* (fourchette), *étame* (puits), *égnak* (vache), *étàme* (seul, e, isolé, e), etc.;
- [o] est noté *o* et il est aussi parfois noté *au* (*Aubame*): *nlo* (tête), *dzope* (ciel), *otsira* (fûmée), *ototo* (motocyclette), *mbolo* (bonjour), *nkoke* (sucre), *okeng* (couteau de table ou de cuisine), *sikolo* (école, études), *étope* (vêtement, habit), *tok* (cuillère), *a zomo* (supporter, être patient), *ékope* (peau), *oyeme*

(langue), *ongoro* (margouillat), *olame* (piège), *owono* (arachide), *oteteghe* (doucement), *otetègne* (étoile), etc.;

- [u] est noté *ou*: *Nkoulou*, *Assoumou*, *Ntoutoume*, à *kou* (tomber, échouer), *aboume* (ventre), *agnou* (bouche), *agnou* (doigt, orteil), *assou* (face, visage), *nsoute* (noir), *nkoute* (nuage), *akoute* (turbulent, e, agité, é), à *gnou* (boire), *mvousse* (dos), etc.

Consonnes

- [b] est noté *b*: *a bi nga* (tirer un fusil), *a bwèlè* (mouiller, se mouiller), *a bi photo* (photographier), *a birane* (gaspiller), *a bwéné agnou* (hurler), *a bwègne* (pouirir), *a bègne* (accoucher), *a bèlè* (naître), *bidzi* (nourriture), *a bo* (être, faire) *a bo nseme/mi nseme* (pécher), etc.;
- [ʃ] est noté *ch*: *chechègne* (sexe), à *chèlè* (cacher), *nchiyè* (lézard), *achègne* (cachette), *ochigne* (rivière), à *chii* (verser, renverser un liquide), à *chi* (bondir, apparaître, sauter), etc.;
- [d] est noté *d*: à *dourou* (conduire, arracher), à *doumane* (plonger), à *dômelô* (clouer), à *doù* (tremper), à *dzeme* (danser), à *djié* (commander, diriger), à *damane* (sauter), à *dzik* (se brûler), *dzop* (ciel), à *dzighi* (se brûler), à *dzime* (éteindre), etc.;
- [f] est noté *f*: *fame* (homme), *fa* (machette), *fope* (bavard, e), *fek* (solution), *flassi* (français, Français, e), *Fala* (France), *felevesse* (éclair), *fefogho* (papaye), à *fite* (appuyer), à *fighi* (boîter), à *fouanghane* (bouger), à *fouk* (remuer, envenimer), à *fogho* (secouer, déplacer), à *foughoulou* (envenimer), *fiè* (avocat), à *fumi* (essuyer), à *fasse* (réfléchir, vérifier), à *fene* (tresser), etc.;
- [gb] est noté *gw*: *Ngwa*, *ngwéngweme* (lion), etc.;
- [j], est noté *y*: *Eyang*, *Oyone*, *Oyane*, *a yo* (vomir), *a yà* (sucrer), *a ya* (couvrir, voiler), *a yo* (vomir), *a yoo* (sauter), *a yeghelane* (prier), *a yéghé* (étudier), à *yeghele* (enseigner), à *yame* (préparer), à *yope* (pêcher), à *yi* (pleurer), à *yiri ossou* (avancer), à *yiri amvousse* (reculer), à *yaghane* (payer, dire aurevoir), *a yemme* (durcir, être fort), *a yeme* (connaître, savoir), *a yebe* (répondre, accepter), *a yiè* (chanter); etc.;
- [k] est noté *c* et *k*: *Haut-Como* (Département administratif), *a kobo* (parler), *a kàme* (s'étonner), *a kame* (défendre), *a kome* (réparer, arranger), *kabane* (mouton), à *kik* (couper, circoncire); *kilisse* (fil à coudre), *ékoutabong* (coude), à *koro* (quitter, partir de), *kuigne* (rat palmiste), à *koupe* (verser, renverser un liquide), *kou* (tortue), *a kuyè* (tousser), *kà* (pangolin), *a kot* (maigrir, tarir), etc.;
- [kp] est noté *kw*: *Okwè* (qui s'écrit et se prononce de plus en plus *Okoué*, *ékwa* (sac), *nkwèrè* (matchette), *okwa* (épervier), *akwa* (chasse mouche, balai

traditionnel), à *kwogho* 'se moquer), *kwogho* (moqueur, - se enquiquêteur, - se), etc.;

- [l] est noté *l*: *a lang* (lire), *lama* (lampe), *a lumi* (se faire prendre au piège ou l'hameçon), *a lote* (passer, dépasser), *a li* (débrousser), *a lèghè* (exploser), *a lere* (montrer), *a le* (jouer), *a lepe* (guider, conseiller), *a lirane* (accompagner), *a lete* (guérir), *alete* (dur, solide), etc.;
- [m] est noté *m*: *mono* (argent), *miyeme* (espace), *mintègne* (douleur), *mouane* (enfant), *missonssôbô* (cime), *minengha* (femme), *metoua* (voiture, automobile), *messoughou* (benjamin, e), etc.;
- [mb] dans *Mba*, *mbolo* (bonjour), *mbong* (manio), *mbeng* (beau, belle, jolie,e), *mbiéne* (mollet), *mbi* (course), *mbî* (terrier), *mbègne* (porte), *mbot* (homme, personne), *mbana* (message, nouvelle), *mboko* (un), etc.;
- [mf] *mfoume* (blanc, blanche: couleur), *mfoukowono* (sauce d'arachide), *mfouk* (quelqu'un, e qui envenime) *mfeme* (pain), *mfime* (mur), *mfiang* (couette), etc.;
- [mv] dans *Mvé*, *mvèt* (cithare, épopée fang), *mvépe* (hérisson), *mvang* (vacin), *mveng* (pluie), *mvame* (grand-parent), *mvouane* (poire); etc.;
- [ɲ] est noté *gn*: *gnou* (corps, couleur), *a gning* (vivre), *gnine* (pou), *égning* (vie), *gno* (serpent), *gniè* (mère), *gnië* (limite), *essègne* (travail), *essègne* (paille), *abègne* (sein), *abègne* (cuisse), à *wuègne* (rire, sourire), *luègne* (rire, sourire), *a gni* (rentrer), à *gnou* (boire), *agnou* (bouche), *agnou* (doigt, orteil), *kuègne* (singé), à *wuègne* (rire, sourire), *vuigne* (ami,e), etc.;
- [n] est noté *n*: *neme* (cœur), *nemme* (sorcier, e, vampireux, se), *nene* (gros, sse), *nome* (mari, époux), *nomme* (vieux, vieille, âgé, e), etc.;
- [nd] est noté *nd*: *nda* (maison), *ndzeme* (maboule, distrait, e), *ndouane* (feu), *ndanga* (rat), *a ndanghala* (rebondir, sauter), *ndok* (gourmand, goulu), *ndôk* (sourd, chocolat local), *a ndamane* (abimer), *ndama* (ballon, balle, football), *ndouroumetoua* (chauffeur, automobiliste), *ndzagha* (bois de chauffage), *ndouravion* (pilote), *ndouroutrain* (cheminot), etc.;
- [ng] est noté *ng*: à *nguièbane* (briller, luire), *nguiéme* (queue), *ngôme* (tamtam), *ngui* (gorille), *ngâ* (fusil), *ngà* (épouse), *ngôpe* (miroir, vitre), *ngope* (chaussure), *ngome* (porc épic), *nguenguang* (médecin traditionnel), *ngape* (part), *ngâne* (caïman, conte, récit), etc.;
- [nk] est noté *nk*: *Nkoulou*, *nkoute* (nuage), *nkô* (corde), *nko* (colline, montée, montagne), *nkagha* (iguane), *nkaghala* (colonne vertébrale), *nkou* (tambour), *nkome* (grand et costaud), *nkok* (sucre, canne à sucre), *nkomengâne* (crocodile), etc.;
- [nl] dans *nlo* (tête), *nlongo* (seau), *nlông* (liane, savane), *nsissime* (esprit), *nlong* (ligne), *nlôpe* (canne à pêche), *nlang* (histoire, récit), *nseme* (pécher), etc.;

- [nt] dans *ntanghane* (blanc, blanche: race), *ntiè* (palabre, dispute), *nto* (ainé, e), etc.;
- [y] très rare en fang, est noté *u*: *a lumi* (se faire prendre au piège ou l'hameçon), etc.;
- De nombreuses lettres sont doubles: *Ella*, *Assoumou*, *Abessole*, *essègne* (*travail*), *essiang* (chat), *ossimane* (pensée, idée), *essoulane* (messe), *essawoula* (fourmi), *essigang* (ogre), etc.

Un exemple de traitement d'article d'un dictionnaire bilingue français-fang pourrait donc se présenter comme suit:

- a, A; première lettre de l'alphabet**, e ékang ossou ya e alphabet
/a,α/ nm inv de A à Z.
- à /a/ *prép* (avec mouvement) **aller** ~ **Libreville** a ke Beyok; **aller** ~ **la maison** a ke a nda (lire *a ka nda*); ~ (dans le temps) ~ **10 ans** mi mbou awome; ~ **la saison sèche** o oyone; (dans une description); **le garçon aux cheveux roux** e mone fame ya bi évele bi essi; (avec être) **je suis** ~ **vous tout de suite** ma zou abong ndi; **c'est** ~ **qui de jouer?** za a le (lire *zaa le*)? **c'est** ~ **toi** wa o ne ya egne (lire *wo na gne*); (marque l'appartenance) ~ **qui est cette montre** za a ne ya e nkolo wui? (lire *za na e nkôlô wui*); **un ami** ~ **moi**; é mvuigne dzame (avec un nombre) **nous l'avons fait** ~ **deux** bi ve bo doebègne; ~ **trois on est serré** bia be la bi ne fefana; **mener 3** ~ **2** a bô ossou ya be la; ~ **1000 francs le kilo** kilo a ne 1000 (toyini).

4.2 Les locuteurs et l'emprunt au français dans la terminologie locale

La main invisible dans les langues ne s'exprime pas peut-être aussi bien que par l'emprunt des mots étrangers par les locuteurs en vue dans l'enrichissement de leurs langues. En paraphrasant la théorie de *la main invisible* en économie, nous pouvons d'abord énoncer comme un des principes de *la main invisible* en rapport avec l'enrichissement de la terminologie des langues, comme suit: *en empruntant un mot étranger et à le remplacer ou pas par un mot local de manière que sa capacité à communiquer ait le plus le plus de valeur possible, le locuteur ne pense qu'à sa propre communication*. L'enrichissement du vocabulaire d'une langue trouve à sa base des milliers de décisions individuelles où chaque locuteur, en empruntant un mot et en le remplaçant ou pas par un mot local, cherche naturellement les moyens de mieux communiquer individuellement. Comme Adam Smith l'a énoncé en économie, on peut dire dans les langues que *ce n'est pas toujours ce qu'il y a de plus mal pour la langue, que cette fin n'entre pour rien dans ses intentions*. On pourrait donc ajouter comme deuxième principe: "C'est tout en ne cherchant que son intérêt personnel dans l'emprunt et par son remplacement ou non, que le locuteur travaille souvent d'une manière bien plus efficace,

comme conduit par une main invisible, pour l'intérêt de la terminologie de la langue, que s'il avait réellement pour but d'y travailler".

Les locuteurs rencontrent souvent des difficultés quand il leur est demandé ex abrupto de trouver des équivalents des mots modernes manquants dans leurs langues locales. Les mots qu'ils proposent dans ce cadre de l'intérêt général de la langue, ont souvent du mal à être plus efficacement intégrés dans le groupe linguistique. C'est le contraire de ceux qui ont été plutôt naturellement élaborés dans un processus de recherche de l'intérêt personnel. Cela signifie que la recherche de la performance individuelle et inconsciente, constitue alors le plus sûr moyen d'œuvrer pour enrichir la terminologie des langues. Il ne serait donc pas recommandé de forcer les locuteurs à trouver des équivalents locaux des mots modernes. Il faut tout simplement collecter les mots qu'ils ont en usage même si certains sont français.

Dans les dictionnaires bilingues monodirectionnels locaux, les personnes en charge de leur confection doivent donc intégrer les mots français que les locuteurs utilisent couramment dans leurs langues locales tels que *Ministre, Député, Sénateur, Ciment, Brique, Carreau, Rideau, Glaçon, Glace, Université, Lycée, Banque, Crédit, Télévision, Dictionnaire, Ordinateur, Téléphone*, etc. Cela ne devrait pas être perçu comme une quelconque anomalie ou une impureté, du moment où ces mots permettent à ces langues locales de remplir pleinement leur fonction de communication. Ces personnes n'ont pas à proposer des termes venant d'eux-mêmes selon leurs connaissances linguistiques et leurs idéologies. Ils n'ont pas également à obliger les locuteurs à produire immédiatement des équivalents locaux en remplacement des mots français. Ce processus va se faire naturellement à partir des milliers de décisions individuelles dans lesquelles les locuteurs, dirigés par *une main invisible*, vont les remplacer ou pas en fonction de la familiarité qu'ils auront avec le concept ou l'objet véhiculé par ces emprunts au français. Ces décisions individuelles seront reversées ensuite dans le cadre de l'intérêt général sans que cela ne soit l'intention de départ des individus.

abonnement /abɔnmã/ *nm* abonnement.
souscrire un ~à sane abonnement; (**carte d'~**)
carte ya abonnement.
addition /adisjɔ/ *nf* addition; (au restaurant)
tang.
additionner /adisjɔne/ *vtr, vpr* à kogolo.
adjectif /adjektif/ *nm* adjectif.

adjutant /adʒɪdã/ *nm* ML adjutant.
aéroport /aeropɔr/ *nm* nsengavion.
pile /pil/ *n f; AUT* (électrique, électronique)
akok; (batterie) batterie.
stade /stad/ *nm*; (terrain de jeu) nsengndama.
stage /staʒ/ *nm* stage.
stagiaire /staʒjɛr/ *nmf* stagiaire.

Conclusion

L'énonciation d'une théorie de *la main invisible* dans les langues semblable à celle de l'économie nous amené à déterminer la place et le rôle respectifs des locuteurs et des experts des langues dans le bon fonctionnement de ces dernières. Les enseignements que nous pouvons tirer de cette réflexion tiennent

lieu de directives dans le traitement des langues locales du Gabon, en vue de leur valorisation et de leur promotion naturelles et durables. Que ce soit aussi bien dans les manifestations orales que dans les manifestations écrites, la langue germe du génie créateur des locuteurs eux-mêmes à qui appartient la langue. Comme mus par *une main invisible*, ces derniers participent à la floraison de leur langue, sans que cela ne soit pas dans leur intention première. Que ce soit en économie ou dans les langues, nous pouvons dire que *la main invisible* traduirait assurément l'adaptabilité exceptionnelle de l'homme: son comportement est régi par des facteurs innés, et par son adaptabilité aux conditions et aux apports internes et externes de son environnement.

Les locuteurs gabonais, comme guidés par *une main invisible*, ont donc déjà posé les jalons d'une approche systématique de la mise en forme, de la valorisation et de la promotion de leurs langues locales et ce particulièrement, dans les aspects liés à l'écriture et au développement de la terminologie moderne. Il ne reste plus aux experts qu'à extraire de ces données les usages récurrents pour en dégager les formes normées d'écriture et de la terminologie locales. Ils n'ont pas à se limiter et à s'auto-satisfaire des propositions scientifiques qui selon les théories linguistiques, consigneraient de façon cohérente et exacte les langues locales du Gabon et procéderaient à des créations lexicales comme processus de développement de leur terminologie. Ils n'ont pas aussi, pour certains, à se cacher maladroitement derrière des intentions nationalistes et idéologiques pour mener un combat d'arrière-garde et trouver le rejet du français comme mode de fonctionnement et de développement de ces langues.

Comme en économie, les mécanismes naturels des locuteurs dans la production orale et écrite et qui participent au fonctionnement de base des langues valent bien toutes les théories. C'est pour cela qu'en dehors des critères théoriques gouvernant la confection de dictionnaire *le lexicographe va vers ses publics, il en est un et ces derniers semblent sans cesse l'interpeller*, cf. Kalonji (1993: 42). Il les implique dans la mesure où l'élaboration de l'écriture et du vocabulaire qui seront consignés dans le dictionnaire est prédéterminée par une certaine réalité et idéalité qu'ils ont de leur langue. Prinsloo (2014) résume bien cet état de fait en disant que *le lexicographe ne devrait pas créer les mots, l'alphabet ou l'orthographe d'une langue. Son traitement dictionnaire devrait refléter leur pratique et leur usage existants. Pour ce faire, la fréquence d'usage est la démarche idéale pour atteindre cet objectif.*

Notes

1. L'administration n'étant pas orale mais écrite, le besoin de représenter les manifestations écrites des langues locales répondait aux soucis de mener à bien les activités politiques, administratives et économiques qui rentraient dans le cadre du développement du Gabon en général.
2. Nous citons Ndong Ngoua, A. L'officialisation des langues africaines: Une question de volonté politique. *Revue semestrielle de l'Institut de Recherche en Sciences Humaines (CENAREST-Gabon)* 11-12: 21-30.

3. Raponda-Walker, A. (Éd.). 1993. *Dictionnaire étymologique des noms propres gabonais*.
4. Raponda-Walker, A. et R. Sillans. 1961. *Les plantes utiles du Gabon: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
5. D'autres mots proviennent également des langues africaines dont les communautés se sont durablement installées au Gabon. Il s'agit entre autres de l'arabe, de l'éwé, de la langue kwa du Ghana ou du Togo, du lingala Congo-Brazzaville et Congo-Kinshasa. Il y a aussi des langues européennes telles que l'espagnol, le portugais et l'allemand dont les populations ont rencontré celles du Gabon au cours de l'histoire des conquêtes européennes. Il y a évidemment l'anglais dont la diffusion mondiale est sans précédent.
6. États Généraux de l'Éducation. *Projet de plan décennal 2010–2020 en matière d'Enseignement, de Formation et de Recherche*. Libreville, 17–18 Mai 2010.
7. Raponda-Walker, A. (Éd.). 1995. *Rapidolangue. Niveau 1. Méthode d'Apprentissage des langues Nationales. Volume 1. (Fang, inzebi, lembama, omyènè, yipunu)*. Raponda-Walker, A. (Éd.). 1996. *Rapidolangue. Niveau 1. Méthode d'Apprentissage des langues Nationales. Volume 2. (ghetsöghö, ihkota)*. Libreville, Gabon: Fondation Mgr Raponda Walker.

Bibliographie

- Chiss, J.-L. et C. Puech.** 1983. La linguistique et la question de l'écriture: enjeux et débats autour de Saussure et des problématiques structurales. *Langue française* 59: 5-24.
- Diop, C.A.** 1979. *Nations nègres et culture*. Tome 2. Paris: Présence Africaine.
- De Saussure, F.** 1964. L'arbitraire du signe. De Saussure, F., A. Sechehaye, C. Bally et A. Riedlinger (Éds.). 1964. *Cours de linguistique générale*: 98-101. Paris: Payot.
- États Généraux de l'Éducation.** *Projet de plan décennal 2010–2020 en matière d'Enseignement, de Formation et de Recherche*. Libreville, 17–18 Mai 2010.
- Hagège, C.** 1987. *Le français et les siècles*. Paris: Odile Jacob.
- Jakobson, R.** 1963. Linguistique et poétique. Jakobson, R. (Éd.). 1963. *Essais de linguistique générale*: 209-248. Paris: Minuit.
- Kalonji, M.T.Z.** 1993. La lexicographie bilingue en Afrique francophone, l'exemple français-cilubà. Paris: L'Harmattan.
- Calvet, Louis-Jean.** 2004. Si les Romains étaient restés à Rome, on parlerait encore le latin. *La gazette de la presse francophone* 115.
- Madiba, M.R.** 2001. Towards a Model for Terminology Modernisation in the African Languages of South Africa. *Language Matters* 32: 53-77.
- Mojela, V.M.** 2014. *Standardizing What We Don't Speak, and Speaking What We Don't Standardize — The Role of Standardization in the Compilation of Standard Dictionaries in Sesotho sa Leboa*. Communication présentée à la 19e Conférence internationale de l'African Association for Lexicography (AFRILEX), 30 Juin–3 Juillet 2014, Potchefstroom, Afrique du Sud.
- Ndong Ngoua, A.** 2007. L'officialisation des langues africaines: Une question de volonté politique. *Revue semestrielle de l'Institut de Recherche en Sciences Humaines (CENAREST–Gabon)* 11–12: 21-30.
- Prinsloo, D.J.** 2014. A Critical Evaluation of the Paradigm Approach in Sepedi Lemmatization — the Groot Noord-Sotho Woordboek as a Case in Point. *Lexikos* 24: 251-271.

- Raponda-Walker, A. et R. Sillans.** 1961. *Les plantes utiles du Gabon: essai d'inventaire et de concordance des noms vernaculaires et scientifiques des plantes spontanées et introduites du Gabon*. Paris: Lechevalier.
- Raponda-Walker, A. (Éd.).** 1993. *Dictionnaire étymologique des noms propres gabonais*. Versailles: Classiques africains.
- Raponda-Walker, A. (Éd.).** 1995. *Rapidolangue. Niveau 1. Méthode d'Apprentissage des langues Nationales. Volume 1. (Fang, inzebi, lembaama, omyènè, yipunu)*. Libreville, Gabon: Fondation Mgr Raponda Walker.
- Raponda-Walker, A. (Éd.).** 1996. *Rapidolangue. Niveau 1. Méthode d'Apprentissage des langues Nationales. Volume 2. (ghetsöghö, ihkota)*. Libreville, Gabon: Fondation Mgr Raponda Walker.

Creating a German–Basque Electronic Dictionary for German Learners

David Lindemann, *Department of Basque Language and Communications, UPV-EHU University of the Basque Country, Donostia, Spain*
(david.lindemann@ehu.es)

Abstract: In this paper, we introduce the new electronic dictionary project *EuDeLex*, which is currently being worked on at UPV-EHU University of the Basque Country.¹ The introduction addresses the need for and functions of a new electronic dictionary for that language pair, as well as general considerations about bilingual lexicography and German as foreign language (GFL).

The language pair German–Basque, which can be called *less-resourced* or *medium-density*, does not have any lexicographical antecedents that could be updated or adapted. Nevertheless, existing monolingual lexicographical databases and a newly created German–Basque parallel corpus support the editing process of the new dictionary. We explain our workflow in macrostructure and microstructure design and editing, and propose a first iteration of the online user interface and publishing process.

Keywords: BILINGUAL LEXICOGRAPHY, ELECTRONIC DICTIONARIES, BASQUE LANGUAGE, GERMAN AS FOREIGN LANGUAGE, PARALLEL CORPORA, USER INTERFACE, WIKTIONARY

Opsomming: Die samestelling van 'n Duits–Baskiese elektroniese woordeboek vir Duitse aanleerders. In hierdie artikel word *EuDeLex*, die nuwe projek vir 'n elektroniese woordeboek wat tans aan die UPV-EHU Universiteit van die baskiese gebied saamgestel word, bespreek. In die inleiding word gewys op die behoefte aan en funksies van 'n nuwe elektroniese woordeboek vir hierdie taalpaar asook algemene aspekte van tweetalige leksikografie met Duits as vreemde taal.

Die taalpaar Duits–Baskies, waarna verwys kan word as 'n taalpaar met minder hulpmiddele en medium digtheid, het geen leksikografiese voorgangers wat hersien of aangepas kan word nie. Desondanks word die samestellingsproses van 'n nuwe woordeboek ondersteun deur bestaande eentalige leksikografiese databasisse en 'n nuwe Duits–Baskiese parallelkorpus. Die werkswyse word bespreek m.b.t. die ontwerp van die makro- en mikrostruktuur en die redigering, en voorstelle word gemaak vir 'n eerste weergawe van 'n aanlyn koppelvlak en die publikasieproses.

Sleutelwoorde: BASKIESE TAAL, DUIJS AS VREEMDE TAAL, ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSKOPPELVLAK, PARALLELKORPORA, TWEETALIGE LEKSIKOGRAFIE, WIKTIONARY

1. Introduction

Basque is today, together with Spanish, official language for the Spanish state

territories Bizkaia, Gipuzkoa and Araba (*Comunidad Autónoma del País Vasco*) and the northern half of Goi Nafarroa (*Comunidad Foral de Navarra*). In the territories belonging to the French state (Lapurdi, Behe Nafarroa and Zuberoa), Basque lacks official status. In the first three areas mentioned, about 30% of the 1.9 million inhabitants are regarded as active bilinguals (Basque and Spanish), and another 11% are regarded to be passive bilinguals (who understand both, but speak only Spanish). For the entire Basque Country, the figures given for active bilinguals are about 25%, the passive being another 10.5%.² In Bizkaia, Gipuzkoa and Araba, today more than a half of secondary and high school students study partly or entirely in Basque language. A third of the nearly 50,000 students at the public University of the Basque Country opts for Basque as teaching language, and the trend is rising.³

For German as a foreign language, the first text book written in German and Basque, without resorting to Spanish or French and designed for Basque-L1 GFL learners, was published in 2007 (Reuter and Wolff 2007). In the following years, a series of studies about teaching German to Basque-L1 learners appeared (Braun 2010, Reuter 2010, Wolff 2010). German teachers in the Basque Country have all made the same experience: Regardless of the teacher having recommended the use of monolingual dictionaries, learners stick to bilinguals, and lacking a suitable offer for German–Basque, they use Spanish–German dictionaries, which in terms of macro- and microstructure have a lot more to offer to them than the only available German–Basque pocket-size dictionary (Martínez Rubio 2007).^{4/5} In order to fill this gap, it is the author's aim to propose and ultimately to provide an alternative, a German–Basque dictionary for Basque-L1 GFL learners, to be made freely available to all users through a web interface hosted by the University.⁶

1.1 *Editing Software*

By means of modern *Dictionary Writing Systems* (DWS) as, for example, *Tshwane-Lex* (Joffe, De Schryver and Prinsloo 2003, De Schryver and Joffe 2005), it is possible to create *Multifunctional Lexical Databases* (about the concept, Pajzs 2009) based on XML-coded microstructures. Different monofunctional dictionaries, which respond to different dictionary functions (Tarp 1995, Tarp and Bergenholtz 2005) can be designed as derivatives from that database, i.e. dictionaries that suit to one user profile and communicative situation (as, for example, text reception or text production). The display of contents from the same source database could also be adapted on the fly to the users' needs, by fading in and out parts of the microstructure or changing the language in which lexicographical metadata are presented. In the age of electronic lexicography, limitations of space or specialized output format disappear.⁷

1.2 *Desiderata for a bilingual lexicography for learners of GFL*

In addition to considerations about what is to be expected from electronic dic-

tionaries, and what expectations electronic dictionaries are able to fulfill (see, for instance, Kemmer 2010, Tarp 2012), studies about bilingual pedagogical lexicography (recently in Fuertes-Olivera 2010) have been published, as well as studies about electronic lexicography with the GFL learner as targeted user profile (Meliss 2013).

In this section we will list some features by which a bilingual dictionary could be considered suitable for the needs of GFL learners. Apart from general desiderata for bilingual lexicography, we can point out some features that are linked to students' difficulties in the learning process of GFL.

1.2.1 *Desiderata for Framing Structure and Macrostructure*

A bilingual dictionary should contain introductory and explanatory prefaces in both languages. User interfaces (UI) of electronic dictionaries should be able to toggle between both languages as metalanguage, i.e. all instructions apart from the lemma signs, synonyms and translation equivalents should be available in both languages.

In addition to canonical word forms that figure traditionally as lemma signs in dictionaries, a UI should also provide results if the user enters an inflected word form in the search box: It may provide a morphological analysis of the word form and a link to the corresponding lemma. Apart from single word lemma signs, also multi-word expressions such as light verb constructions or idiomatic phrases should be capable to be found in a dictionary.

1.2.2 *Desiderata for the Microstructure*

A bilingual dictionary article should be furnished not only with insightful instructions for word sense (polysemy) disambiguation and a mapping to suitable Translation Equivalents, but also with instructions related to morphology, syntax, and pragmatics. In the case of a dictionary for GFL learners these are the following:

- Inflection morphology paradigm for the German verb, noun or adjective
- For German verbs: auxiliary selection
- Instructions related to valency (argument structure realisation)
- German synonyms
- Frequency data
- Collocates
- Pragmatics (register)
- Example sentences from bilingual corpora

1.3 *Lexicography and Open Source*

For a *high-density* language like German, many printed and electronic resources are available today. Innumerable dictionaries have been compiled by lexicographers, and phonetic, morphological, syntactic and semantic information about the headwords given in those has been edited many times or updated from former editions. A new dictionary that can not take advantage of relevant antecedents and that is not born in the shade of a publishing house that possesses extensive lexicographical databases, nevertheless may start with data obtained from open sources, i.e. sources that are not only freely accessible from an online UI, but also available as complete source files under non-exclusive, non-proprietary licensing terms. For the case of German, today these are primarily the electronic lexicographical databases *openthesaurus* and the German edition of the crowdsourced *Wiktionary*, that are continuously growing and possibly will reach a professional level (Meyer and Gurevych 2010, 2012). Furthermore, for the definition of a German lemma list, the corpus based lemmatized frequency word lists released by the IDS under a Creative Commons licence (see section 2.1) are a suitable starting point.

Generally, lexicographical data from existing databases can be used for a new project in two ways: By importing data into the new database as draft for manual revision (Dictionary Drafting), or by dynamically including data from external sources in the search-result pages of an online dictionary, as described in section 3.

In the case of Basque, a *medium-density* minority language with less than one million speakers and a co-official legal status in parts of the territory it is spoken and a relatively minuscule web size (see Lindemann 2013), noteworthy open resources are available: The source files of *Hiztegi Batua* (Euskaltzaindia 2008), and the Basque WordNet (EusWN, see Pociello, Agirre and Aldezabal 2011), that are both published under Creative Commons licences.

1.4 *A German–Basque literary parallel corpus*

At the University of the Basque Country, a German–Basque Literary Corpus has been created (Sanz Villar 2013, Zubillaga 2013), using the content of 81 digital or digitized literary German originals and their official, direct translations into Basque. In its current version, the German–Basque translation corpus counts about 2 million tokens per language. The sentence alignment (146.000 sentence pairs) had to be revised manually, starting from an automatic alignment at paragraph level. Today, that corpus is the only parallel resource for German–Basque.⁸ The value of parallel corpora for pedagogical lexicography (cf. Bowker 2010) as well as the value of parallel (literary) corpora for lexicography in general are beyond question (cf. Teubert 2002).

This parallel corpus has been imported into the *SketchEngine* (Kilgarriff, Rychly, Smrz and Tugwell 2004), a software system which is capable of dealing

with parallel corpora. The German part was lemmatized and POS-tagged with *TreeTagger* (Schmid 1995), a tool that is built in to the *SketchEngine*. The Basque corpus has been lemmatized with *EusTagger* (Aduriz, Aldezabal, Alegria et al. 1996). As a result of the tagging and lemmatization, the corpus can be queried by lemma, in order to consider all appearances of inflected forms of the lemma in the corpus. This is a desirable feature for both dictionary entry drafting process and for display on dictionary search result pages.

In the dictionary editing process, data from this literary corpus has been particularly useful for German lemma signs denoting abstracts: For some abstract nouns and verbs, considered "hard tasks" for a lexicographer, where data from other sources does not lead to satisfying results, the corpus data provides groups of good Translation Equivalent candidates, each one reflecting a translator's choice in a particular context. Table 1 illustrates this by two German nouns, two often cited examples for "hardly translatable" German abstract nouns:

<i>German lemma sign (counts)</i>	<i>Basque TE from parallel corpus</i>
Gemütlichkeit (4)	goxotasun, patxada, lasaitasun, konfortea
Schadenfreude (10)	(voll Schadenfreude sein) zoritxarraz poztu (Schadenfreude empfinden) maltzur sentitu bozkario gozatze modu bat poz txiki bat alaitasun maltzur besteak umiliatzeko poza poz gaizto (aus Schadenfreude) besteren gaitzak ninduen pozten kalte poz

Table 1: German abstract nouns and Basque TE from literary parallel corpus

Apart from the described function as documentation for the lexicographer in the lexicographical workflow, a display of parallel corpus concordances as part of the search results in online bilingual dictionaries may be worthwhile (first mention in Atkins 1996, Dickens and Salkie 1996), and it is a desired feature for *EuDeLex*.

2. A new German–Basque Electronic Dictionary

EuDeLex is in its first stage of development. A macrostructure and a microstructure for DE>EU have been proposed, dictionary entries have been edited for around 10% of the planned lemma list (4,500), and a preliminary version of the online user interface is being tested.

This first-stage work is being done on the German–Basque side, but there is also the possibility of including a Basque–German part. There may also be occasion to widen the scope of the dictionary and adopt its structure towards other functions than the described.

The *EuDeLex* dictionary articles for German Letter A, already edited, have served as gold standard in the evaluation of six different corpus based and/or lexical knowledge based Bilingual Dictionary Drafting methods, which has been carried out by the author together with a group of computational linguists (Lindemann, Saralegi, San Vicente, Manterola and Nazar 2014). The (semi-)automatically produced bilingual glossaries are evaluated quantitatively and qualitatively. Results show that the described methods can greatly assist the editing of dictionary entries for the remaining 90% of the German lemma list. Furthermore, the set of Bilingual Dictionary Drafting methods described may serve as reference for lexicographical work on other language pairs that starts from scratch.

2.1 *Macrostructure*

It has been proved that the most frequent words are actually the words most frequently looked up by dictionary users; this is true for the top few thousand (De Schryver, Joffe, Joffe and Hillewaert 2006, Wolfer, Koplenig, Meyer and Müller-Spitzer 2014). At the same time, frequency data is useful information for both dictionary editor and user. Therefore, it makes sense to build a lemma list starting from corpus-based frequency lists and to include frequency data in the published dictionary. For German, lemmatized frequency lists based on large reference corpora are available under public licences (IDS 2009). In our workflow, we compare the DeReWo-40.000 frequency word list with human-revised lemma lists found in three editorial dictionaries, and delete, replace (by adapting to a word form used as lemma in our macrostructure) or add lemmata. DeReWo-40.000 contains the whole lemma list regarded as a GFL learner's basic dictionary (*Wortschatz Zertifikat Deutsch als Fremdsprache*, CEFR B2).

After editing dictionary entries for the German letter A, i.e. the first 3453 entries on the (alphabetically ordered) DeReWo-40,000, our adaptation omitted 132 of those 3453 entries from our lemma list. 101 of the non-imported lemmas were not found in the DUDEN, the first secondary source for reference, mostly so-called [semantically] transparent compound nouns. Another 24 were proper names of people or organizations ("AOK") or topo- and hydronyms ("Alster"), which we generally do not import, and 7 others. Another 37 DeReWo entries we have incorporated in a modified form (mainly led by the form listed in the DUDEN), such as "Abbrucharbeiten" instead of "Abbrucharbeit" or "aufrütteln" instead of "aufrütteln." In the first edition of *EuDeLex*, we restrict the published dictionary articles to nouns, verbs, adverbs, and adjectives (see section 2.2.1), as function words in our context bear special difficulties that need further research.

For a definition of homonymy we follow Kempcke's (2001: 67) criterion of

morphosyntactic disparity as a necessary condition: If homograph lemma signs follow different inflection morphology patterns or if they are morphological derivatives with different origins, they will be considered as homonyms, which appear in separate dictionary articles. We assume that it is helpful for students to distinguish words with different inflectional morphologies, which in the event they must learn separately.

In table 2, we list absolute amounts and percentages of nouns, verb infinitives, verb participles, adjectives, and adverbs in *EuDeLex* German Letter A (DeReWo Letter A edited as described above), DeReWo A-Z tagged with RFTagger (Schmid and Laws 2008), and the intersection of DeReWo and the German Wiktionary (wordlist and POS-tags parsed from de.wiktionary).

	EuDeLex DE Letter A		DeReWo A RFTagger		DeReWo A-Z RFTagger		DeReWo \cap de.wiktionary	
<i>Lemmata</i>	3,614		3,453		40,000		22,404	
Nouns	1,963	54.32%	1,813	52.51%	25,529	63.82%	14,017	62.56%
Verb Infinitive	924	25.57%	878	25.43%	5,813	14.53%	3,203	14.30%
Verb Participle	268	7.42%	175	5.07%	819	2.05%	1,003	4.48%
Adjectives	413	11.43%	411	11.90%	5,243	13.11%	3,418	15.26%
Adverbs	150	4.15%	60	1.74%	579	1.45%	585	2.61%
Others	0	0.00%	116	3.36%	2,017	5.04%	461	2.06%
<i>Total</i>	3,718	102.88%	3,453	100.00%	40,000	100.00%	22,687	101.26%

Table 2: Word classes in the German lemmalist

The above figures show a disproportionately high rate of verbs in German letter A, which is due to the presence of three prepositions, and consequently of verbs formed by a chain of preposition plus verb (such as *abfahren*, *ankommen*, *auffallen*). The disparity between verb participle counts in *EuDeLex A* and *DeReWo A RFTagger* can be explained by the fact that, in *EuDeLex*, participial adjectives are tagged as both adjectives and verb participles (together with a link to the corresponding verb infinitive). The higher number of adverbs in *EuDeLex* is due to the disambiguation between adjectives and adjectives with adverbial use (see section 2.2.4), and a double tagging in cases where both uses are found in German. As *EuDeLex* POS-tags are manually set, these figures also demonstrate the high grade of trustworthiness of the *RFTagger* tool.

For Basque, the method for lemma list building we are aiming at is the following: The Basque Language Academy *Euskaltzaindia* provides a dictionary with a 55.000 entry corpus-based lemma list under a public licence (*Euskaltzaindia*

2008), each entry of which has been revised and approved by the Academy's lexicographical board. A comparison of that list with lemmatized frequency word lists extracted from Basque Web Corpora is in course, in order to define a core lemma list as starting point for *EuDeLex* EU>DE (Lindemann and San Vicente in prep., for a survey of Basque corpora see Leturia 2012).

2.2 *EuDeLex* Microstructure

2.2.1 *Structure of the editorial dictionary articles*

The aim to provide a dictionary in the reasonably near future lies in conflict with limited human and other resources in dictionary entry editing. For this reason, we don't hesitate to take pragmatic decisions in prioritizing the lexicographical working agenda. The aim of this dictionary in its first version we define as to provide useful word sense disambiguation (polysemy discrimination) information to Basque-L1 GFL learners and translators. Accordingly, the Basque equivalents provided in the first version of the German–Basque part of *EuDeLex* will be furnished with additional information when that seems useful for polysemy discrimination: Possible information to be included is specification of semantic domain, synonyms and/or usage examples, as shown in Fig. 1:

abdecken Derewo 4349 *trennbar*
Verb Transitiv -haben ▶ **1 desestali** *das Dach abdecken* **teilatua kendu**
2 (mahaia) jaso *Der Kellner hat den Tisch abgedeckt* **Zerbitzariak**
mahaia jaso du **3 estali** *Im Winter deckt sie das Beet ab* **Neguan**
baratzea estaltzen du; Die Antenne deckt die ganze Region ab
Antenak eskualde osoa estaltzen du

Absatz Derewo 3653 ~es, Absätze
Substantiv m. ▶ **1 (oinetakoa) takoi; orpo** **2 (idatzia) paragrafo;**
lerroalde **3 HANDEL salmenta**

Abstufung Derewo 26176
Substantiv f. s. Verbinf. abstufen ▶ **1 mailaketa** **2 = Herabstufung**
mailaz jaiste **3 = Nuance ñabardura**

Figure 1: Excerpts from *EuDeLex*

After the first version is edited, we plan to provide all dictionary entries with such information. Merely with regard to the domain specification there is a claim for completeness from the first version on, in order to provide domain related glossaries as derivatives of *EuDeLex*.⁹

2.2.2 *German nouns in dictionary articles*

Following the criteria described in section 2.1, we provide, as seen in Fig. 2, the information on the German noun "Ausdruck" in two different dictionary entries

and add to both information about inflection morphology (genitive sg. and nominative pl. forms) together with the grammatical gender.¹⁰

Ausdruck¹ Derewo 1918 ~s, Ausdrücke
Substantiv m. s. Verbinf. ausdrücken ▶ 1 adierazpen 2 termino; esamolde; esapide
 3 ezaugarri; adierazgarri 4 adierazkortasun; adierazgarritasun

Ausdruck² Derewo 1918 ~s, Ausdrücke
Substantiv m. s. Verbinf. ausdrücken ▶ 1 inprimatze; inprimaketa

Figure 2: Excerpts from *EuDeLex*

Below this first syntactic categorization, the polysemy follows in numbered word sense sections. In distinguishing word senses, we try to take into account possible asymmetric lexicalization (about the concept, cf. Hartmann 2007: 33) in German and Basque. Lexical asymmetry of a group of hyper- and hyponyms can be exemplified as illustrated in Figure 3. When there is no suitable bilingual dictionary available for the language pair a dictionary user wants to find information on, asymmetry schemes like this show why using a third language's dictionaries as "pivot" or "bridge" may mislead, why the polysemy of a German word must be mapped lexicographically to a target language's lexical units in a specific way:

DE		EU		EN	ES	
Holz	Brennholz, Feuerholz	egur	su-egur	firewood	leña	
	Bauholz		zur	lumber, timber	madera	
	Holz	egur, zur		wood		
	Gehölz	Wald	ohian, baso		woods, forest	
	Forst	baso (ustiatu), zuhaizti	forest, woodland			

Figure 3: Asymmetric lexicalization

The above mentioned German terms therefore should be paired with the following equivalents (see table 3):

Holz	1. egur, zur; 2. ohian, baso
Gehölz	baso, zuhaizti
Wald	ohian, baso
Forst	baso, zuhaizti (ustiatu)
Brennholz	su-egur

Feuerholz	su-egur
Bauholz	(eraikuntza) zur

Table 3: German "Holz" and hyponyms, and Basque equivalents

2.2.3 German Verbs in dictionary articles

Following microstructures established in some GFL dictionaries (Langenscheidt pocket Dictionaries, PONS dictionary GFL) we organize the verb entry first and foremost syntactically, i.e. below a syntactical entity (transitive, intransitive, reflexive/reciprocal, non-personal use, see Fig. 3 below).¹¹

Some arguments in favour of such syntactic element ordering may be summarized as follows (cf. also Marelllo 2010, Dentschewa 2006: 124f):

- It helps as a first orientation within longer dictionary entries
- In text reception, it is a strategy of advanced GFL learners to identify the verb (the meaning of which they possibly don't know) and its arguments as syntactic entities (subject, accusative object, dative object etc.) and then to proceed to semantics.
- In GFL production, it is often not the meaning but the syntactic properties of a word that dictionary users want to be sure about: Can I use this verb as a transitive? Which auxiliary is selected by that verb in an intransitive sentence?

Criticism of syntactic ordering is based on questioning the ability of dictionary users to deal with basic grammatical concepts such as "transitivity". An alternative could be a labelling as proposed in PONS: "Mit OBJ" (with object), "Ohne OBJ" (without object), "Mit Sich" (with "sich").¹²

abstimmen Derewo 2628 trennbar
 I **Verb Intransitiv** -haben ▶ 1 [über etw.] **bozkatu**
 II **Verb Transitiv** -haben ▶ 1 [auf etw.] **egokitu** 2 [mit jdm.] **adostu; hitzartu**
 III **Verb + 'sich'** -haben ▶ 1 [mit jdm.] **ados jarri**

Figure 4: Excerpt from *EuDeLex*

As support for a usually permanent problem in the learning process of GFL, the verb and its auxiliary selection in perfect tense, the appropriate auxiliary verb is indicated together with the syntactical tag. For a lexicographical illustration of valency, we use the so-called "valence formula" (Wolski and

Cyffka 2011: 12). By this formula, the user gets clues about correct connection of actants with prepositions not only for text production, but also in text reception, as in cases like in Fig. 4 the preposition reveals the necessary for identifying the German verb's word sense. Following the above-defined initial target of this dictionary, in the first version we don't provide translations of the German valency formula. A future upgrade of the existing microstructure targeting German-speaking learners of Basque could look like the following:

bedanken Derewo 5233 ohne ge
 Verb + 'sich' +haben ▶ 1 [bei jdm. für etw.] [nbi. zbtengatik] **eskerrak eman**

Figure 5: Excerpt from *EuDeLex*

By means of valency formula we also specify syntactic properties of verbs as refinement of the rough classification in the above-described sections, primarily ditransitivity ("somebody something"). Reflexive and reciprocal structures with a reflexive pronoun in dative case, as opposed to those with an accusative pronoun, are not grouped in the "with SICH" section, but are marked as transitive verb with the additional tag [*jmdm. etw.*] ("somebody something"), as shown in Fig. 6:

ausleihen Derewo 7354 unregelmäßig, trennbar
 Verb Transitiv +haben ▶ 1 [*jmdm. etw.*] **maileguz utzi; mailegatu**
 2 [*(sich (Dat.)) etw.*] **maileguz hartu; mailegatu**

Figure 6: Excerpt from *EuDeLex*

2.2.4 Adjectives and adverbs in dictionary articles

Regarding adjectives and adverbs, we must keep in mind morphosyntactic symmetries and asymmetries in both German and Basque. As the examples in Fig. 6 show, German adjectives in attributive or adverbial usage require morphologically or lexically different Basque equivalents; the attributive adjectives must be inflected in both German and Basque (being part of a nominal or determinative syntagma), while the adverbial is used without inflection in both languages (which is characteristic for an adverbial syntagma). In Basque, this category change in use of certain lexemes between adjective or adverb requires certain affixes (e.g. *-ki*, *-to*, *-ka* to obtain adverbs, and *-ko* to obtain adjectives), which also applies to some German lexemes (cf. *allein_{ADV}* vs. *alleinig_{ADJ}*). The Basque equivalents of German adjectives and adverbs can also appear as postposition syntagma (*jendetasun_{NOM}-ez_{POSTP}*, "with respectability").

anständig Derewo 8054
I **Adjektiv** ▶ 1 **zintzo**
II **Adjektiv adverbial** ▶ 1 **zintzo; zintzoki; jendetasunez**

anteilig Derewo 20466
I **Adjektiv** ▶ 1 **zatikako; proportzional**
II **Adjektiv adverbial** ▶ 1 **zatika; proportzionalki**

arg Derewo 3262
I **Adjektiv** ▶ 1 **latz; gaitz** 2 *veraltend* **gaizto; maltzur**
II **Adjektiv adverbial** ▶ 1 *ugs. südd.* **oso**

Figure 7: Excerpts from *EuDeLex*

In various cases, lexemes of both languages only appear in one of the two categories, i.e. they can not be modified by affixes and made use of in the other category. As in this context either German nor Basque has standard rules that would apply universally, GFL learners would ask their dictionary for proper uses of the doubtful lexeme as adjective or adverb (see Fig. 8).¹³

anders Derewo 596
I **Adverb** ▶ 1 **ezberdin; bestela; beste era batean**
II **Adverb präd./att.** ▶ 1 **ezberdin; bestelako; beste era bateko**

Figure 8: Excerpts from *EuDeLex*

Consequently, the entry for German adverbs in use as adverbial or as a predicate noun is organized in a similar way: A German adverb in predicative use ("Jazz ist anders_{ADV}" ("Jazz is different")) does not allow inflection; the equivalent use in Basque requires an adjective, i.e. it necessarily requires inflection ("Jazz-a ezberdin_{ADJ}-a_{DET}/bestelako_{ADJ}-a_{DET} da") The same is true for an attributive use of German Adverbs "die_{DET} Sitzung_{NOM} gestern_{ADV}" vs. "[atz_{ADV}-ko_{POSP} batzar_{NOM}]-ra_{DET}" ("the meeting yesterday/yesterday's meeting"). In Fig. 8, for German adverbs, on the one side non-inflectable equivalents are given for an adverbial use, and inflectable equivalents for a predicative or attributive use, on the other.

3. *EuDeLex*: Online-Publishing

As of May 2013 *EuDeLex* is online.¹⁴ The UI (see Fig. 8 below) is based on a single PHP script. On top of the home page, the user is prompted for a German headword or word form to look for. Starting with the fourth character that is typed, a search routine compares the query string with the *EuDeLex* lemma list, and responds to the user with automatic suggestions to complete a headword. Placed before the search box, switch buttons are provided for selecting the

desired metalanguage so the searches performed will lead to result displays with instructions in German or Basque, i.e. the displayed data from *EuDeLex* concerning syntax, pragmatics and domain, as well as section titles and pieces of advice given in all parts of the result page.

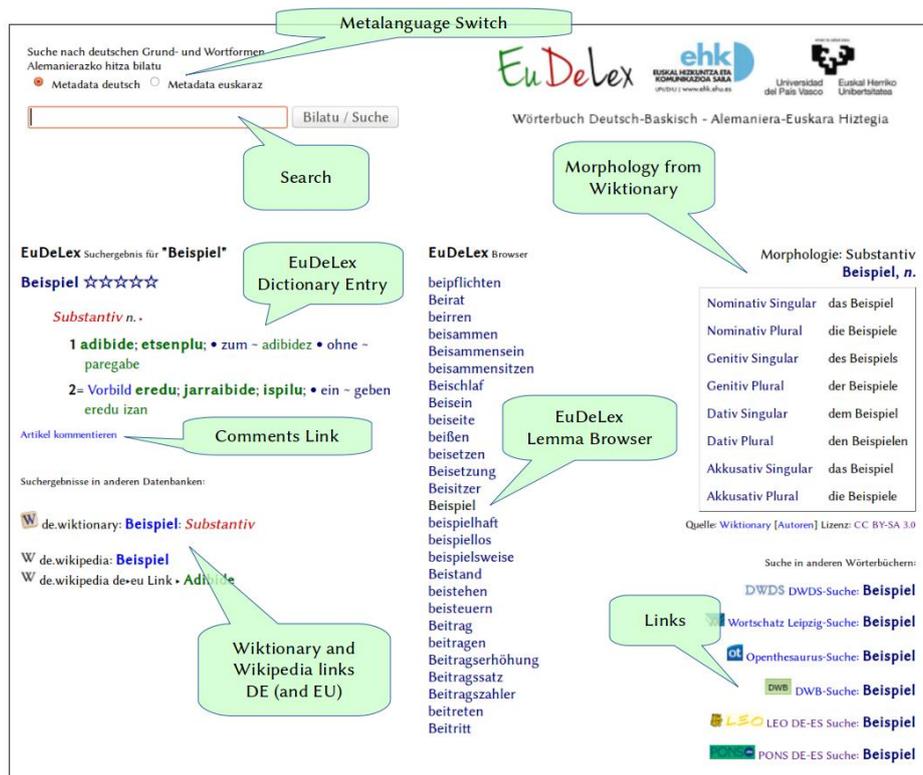


Figure 9: *EuDeLex* User Interface with annotations

The search result frame is organized in three columns:

1. *EuDeLex* dictionary entry or entries (in case of homographs¹⁵) and search results from the German editions of *Wiktionary* and *Wikipedia*, from where the following is retrieved on the fly¹⁶ and, if found, shown to the user: Links to matching pages, and Basque translation links found in those sources.
2. *EuDelex* Lemma browser: Direct neighbours on the *EuDeLex* lemma list are listed as links that lead to the corresponding entry.
3. Morphological information about the German lemma sign, retrieved on the fly from *Wiktionary*, and, if found, reproduced as table. Below, auto-

matically generated hyperlinks to four German monolingual dictionary websites, containing the query string.¹⁷

The queries to the local database (*EuDeLex*) and to *Wikimedia* servers function for their own, i.e. the script also generates results in case the query has not lead to any entry in *EuDeLex*¹⁸. As all data from *Wiktionary* and *Wikipedia* is retrieved on the fly, it will always reflect the latest version of these sources. As for today, for more than a half of the planned dictionary lemma list there already exist extensive articles in *Wiktionary*, and in the case of verbs, nouns and adjectives we can extract and display Wiktionary's tables of inflectional and morphological information (cf. Table 2).

If in column (1) just one *EuDeLex* headword is listed, the user obtains the possibility to comment on the article. For this purpose, a dialogue is opened when clicking on the comment link.

Basque Translation Equivalents in the *EuDeLex* dictionary article, provided they exist as Basque lemma¹⁹, lead by clicking on them to the launch of the Basque–German search script with the selected equivalent as query string, that will also show redirects to Basque dictionary platforms.

The software tools and scripts involved in the publishing are:

- *EuDeLex* data is exported by the *TshwaneLex Dictionary Writing System* to XML format.
- A *perl* script transforms this XML into a MySQL table to be installed on the webserver.
- The above-described UI for the search direction DE>EU is based on one PHP script, that performs the queries from *EuDeLex*-MySQL-database and retrieves data from the *Wikimedia* API and displays the results.
- Another PHP script is planned for search direction EU>DE. In the current preliminary version this script offers automatically generated hyperlinks to Basque dictionary websites containing the query string.

3. Conclusions

By the lexicographical factory report on hand we also hope to provide ideas to projects on other medium-density language pairs that do not have adaptable predecessors, requiring that a dictionary be started from scratch. It has been pointed out that by using modern specialised software it is possible to create multifunctional lexicographical databases, the derivatives of which may serve different dictionary functions. In the event that monolingual corpora or even parallel translation corpora are available for the language pair, the data these sources provide can be gainfully taken into account, alongside data from other sources like existing monolingual dictionaries and dictionaries of third languages. We have also shown our attempt to provide an online-platform for the

dictionary by means of a simple PHP script, as well as to exploit existing and free-licensed lexical data from *Wikimedia* servers, in order to enrich the dictionary search result displays according to the needs of the user profile we target.

4. Future Work

Having defined a Basque lemma list for EU>DE (Lindemann and San Vicente in prep.), and a microstructure for EU>DE dictionary articles, those will be edited, starting from a draft obtained by reversing the DE>EU part of *EuDeLex*. Furthermore, we will implement a parallel corpus concordance line generator in both DE>EU and EU>DE PHP scripts, so that search result displays will be complemented with translated usage examples. An upgrade of the existing German–Basque parallel corpus is also planned.

5. Notes

1. The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465, and from project IT665-13, funded by the Basque Government. Funding is gratefully acknowledged.
2. Source: EUSTAT statistics agency, Eusko Jaurlaritz 2011.
3. Source: UPV-EHU University of the Basque Country 2012.
4. For a survey of Basque Lexicography, see Azkarate (1991, 2014); For German–Basque lexicographical work from 1817 until today, Lindemann (2014).
5. At the moment, we lack any profound investigation in dictionary use (*Wörterbuchbenutzungsforschung*) in this context.
6. The online UI is available at <http://www.ehu.es/eudelex>.
7. See De Schryver (2003) for a commented bibliography on lexicography in the electronic age, and Tarp (2012).
8. An upgrade of this corpus as well as the creation of a German–Basque Bible corpus are in development (for the Bible as parallel corpus, see Resnik et al. (1999)).
9. These domain-specific glossaries could be contrasted with other German glossaries of the same domains, in order to detect gaps and to obtain more "term-wordsenses" or lemma candidates for the main database.
10. As seen in Section 3, we also add morphological information imported from Wiktionary to *EuDeLex* search results.
11. A monolingual and bilingual dictionary entry structure often reflects the polysemy of the headword first and foremost. This implies to include the description of syntactical features of the headword like transitivity in "gram-groups" as child-elements of wordsense-level (e.g. TEI-Standard (Burnard and Sperberg-McQueen 2007)).
12. "Object" here refers only a direct object (accusative), and not to an indirect object (dative), which can be misleading. The term "transitive" seems to us a more appropriate, since transi-

tivity as syntactic property in Basque is a well-known phenomenon: A subject in a transitive clause is set in ergative case, while a subject in an intransitive clause stays absolutive.

13. Once the lexicographer establishes this distinction when editing the entry, a correct mapping to Basque equivalents is also provided, although a Basque-L1 dictionary user might have no need for it. In this sense it is an anticipation that applies to the needs of Basque-L2 users, not in the focus of *EuDeLex* version 1.
14. <http://www.ehu.es/eudelex>.
15. For result display, case sensitivity is not taken into account. If the search result contains more than one homograph headword, the user is asked in column (3) to choose one by clicking on the headword, in order to get morphological information.
16. *Wikipedia* and *Wiktionary* API are queried for the existence of a homograph page title or redirect page title. No disambiguation of homographs or homonyms is attempted.
17. If those URL do lead to any existing dictionary entry on those platforms is not verified.
18. Consequently, Basque GFL learners might start to use this UI, although *EuDeLex* still does not cover a significant part of the German lemmalist.
19. Until the Basque–German part of the dictionary is published, this *script* verifies if the Basque TE exists on a Basque lemma list (HB, Euskaltzaindia 2008) and, in the positive case, provides a link.

6. Bibliography

- Aduriz, I., I. Aldezabal, I. Alegria, X. Artola, N. Ezeiza and R. Urizar.** 1996. EUSLEM: A Lemmatizer/Tagger for Basque. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C.R. Papmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 17-26. Gothenburg: Department of Swedish, Göteborg University.
- Atkins, B.T.S.** 1996. Bilingual Dictionaries: Past, Present and Future. Gellerstam, M., J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C.R. Papmehl (Eds.). 1996. *Euralex '96 Proceedings I-II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 515-546. Gothenburg: Department of Swedish, Göteborg University.
- Azkarate, M.** 1991. Basque Lexicography. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1991. *Wörterbücher: Ein internationales Handbuch zur Lexikographie*. Vol. 3: 2371-2375. Berlin: De Gruyter.
- Azkarate, M.** 2014. On-Line Dictionaries of a Minority Language in a Multilingual Society. *XVIth Forum for Iberian studies Cultural and Linguistic Diversity in the Iberian Peninsula*. University of Oxford, Oxford.
- Bowker, L.** 2010. The Contribution of Corpus Linguistics to the Development of Specialised Dictionaries for Learners. Fuertes-Olivera, P. (Ed.). 2010: 155-168.
- Braun, S.** 2010. Die Rolle der Muttersprache im Deutschunterricht in zweisprachigen Gebieten am Beispiel des Baskenlandes. Jarillot, C. (Ed.). 2010. *Bestandsaufnahme der Germanistik in Spanien: Kulturtransfer und methodologische Erneuerung*: 25-36. Bern: Peter Lang.
- Burnard, L. and C.M. Sperberg-McQueen.** 2007. TEI P5: Guidelines for Electronic Text Encoding and Interchange. *TEI Text Encoding Initiative*. Available: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (Accessed 22 July 2014).

- Dentschewa, E.** 2006. DaF-Wörterbücher im Vergleich: Ein Plädoyer für 'Strukturformeln'. Dimova, A., V. Jesenšek and P. Petkov (Eds.). 2006. *Zweisprachige Lexikographie und Deutsch als Fremdsprache: drittes Internationales Kolloquium zur Lexikographie und Wörterbuchforschung, Konstantin Preslavski-Universität Schumen, 23.–24. Oktober 2005*: 49-58. Germanistische Linguistik 113–128. Hildesheim/New York: G. Olms.
- De Schryver, G.-M.** 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver, G.-M. and D. Joffe.** 2005. One Database, Many Dictionaries–Varying Co(n)text with the Dictionary Application TshwaneLex. Ooi, V.B.Y., A. Pakir, I. Talib, L. Tan, P.K.W. Tan and Y.Y. Tan (Eds.). 2005. *Words in Asian Cultural Contexts, Proceedings of the Fourth Asialex Conference, 1–3 June 2005, M Hotel, Singapore*: 54-59. Singapore: Department of English Language and Literature & Asia Research Institute, National University of Singapore.
- De Schryver, G.-M., D. Joffe, P. Joffe and S. Hillewaert.** 2006. Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. *Lexikos* 16: 67-83.
- Dickens, A. and R. Salkie.** 1996. Comparing Bilingual Dictionaries with a Parallel Corpus. Gellerstam, M., J. Järborg, S.G. Malmgren, K. Norén, L. Rogström and C.R. Papmehl (Eds.). 1996. *Euralex '96 Proceedings I–II, Papers Submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*: 551-559. Gothenburg: Department of Swedish, Göteborg University.
- Euskaltzaindia.** 2008. *Hiztegi Batua*. Donostia: Elkar.
- Fuertes-Olivera, P. (Ed).** 2010. *Specialised Dictionaries for Learners*. Lexicographica Series Maior 136. Berlin: De Gruyter.
- Hartmann, R.R.K.** 2007. The Not so Harmless Drudgery of Finding Translation Equivalents. Hartmann, R.R.K. (Ed.). 2007. *Interlingual Lexicography*: 30-37. Lexicographica Series Maior 133. Berlin: De Gruyter.
- IDS.** 2009. Korpusbasierte Wortgrundformenliste DEREW0, v-4000g-2009-12-31-0.1, mit Benutzerdokumentation. Institut für Deutsche Sprache, Programmbereich Korpuslinguistik. Available: <http://www.ids-mannheim.de/kl/projekte/methoden/derewo.html> (Accessed 22 July 2014).
- Joffe, D., G.-M. de Schryver and D.J. Prinsloo.** 2003. Computational Features of the Dictionary Application 'TshwaneLex'. *Southern African Linguistics and Applied Language Studies* 21(4) [Special issue on 'Human Language Technology in South Africa: Resources and Applications']: 239-250.
- Kemmer, K.** 2010. *Onlinewörterbücher in Der Wörterbuchkritik. OPAL 2*.
- Kempcke, G.** 2001. Polysemie oder Homonymie? Zur Praxis der Bedeutungsgliederung in den Wörterbuchartikeln synchronischer einsprachiger Wörterbücher der deutschen Sprache. *Lexicographica* 17: 61-68.
- Kilgarriff, A., P. Rychly, P. Smrz and D. Tugwell.** 2004. The Sketch Engine. Williams, J. and S. Vessier (Eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*: 105-116. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.
- Leturia, I.** 2012. Evaluating Different Methods for Automatically Collecting Large General Corpora for Basque from the Web. Kay, M. and C. Boitet (Eds.). 2012. *Proceedings of the 24th International Conference on Computational Linguistics, 8–15 December 2012, Mumbai, India (COLING 2012: Technical Papers)*: 1553-1570. Mumbai, India: The COLING 2012 Organizing Committee.

- Lindemann, D.** 2013. Bilingual Lexicography and Corpus Methods. The Example of German–Basque as Language Pair. *Procedia — Social and Behavioral Sciences* 95: 249-257.
- Lindemann, D.** 2014. Zweisprachige Lexikographie des Sprachenpaares Deutsch-Baskisch. Domínguez Vázquez, M.J., F. Mollica and M. Nied (Eds.). 2014. *Zweisprachige Lexikographie zwischen Translation und Didaktik*: 259-281. *Lexicographica Series Maior* 147. Berlin/Boston: De Gruyter.
- Lindemann, D., X. Saralegi, I. San Vicente, I. Manterola and R. Nazar.** 2014. Bilingual Dictionary Drafting. The Example of German–Basque, a Medium-Density Language Pair. Abel, A., C. Vettori and N. Ralli. (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15–19, 2014*: 563-576. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Lindemann, D. and I. San Vicente.** In preparation. *Euskarazko Maiztasun Lemategia Gaurko Teknologien Ikuspuntutik*. Bilbao: UPV-EHU.
- Marelló, C.** 2010. Verbos con construcciones tanto transitivas como intransitivas y/o pronominales en los diccionarios monolingües y bilingües italianos y españoles. Castillo Carballo, M.A. and J.M. García Platero (Eds.). 2010. *La Lexicografía en su dimensión teórica*: 411-434. Malaga: University of Malaga.
- Martínez Rubio, E.** 2007. *Euskara alemana hiztegia*. Donostia: Elkar.
- Meliss, M.** 2013. Online-Lexikographie im DaF-Bereich: Eine erste kritische Annäherung; Bestandsaufnahme — Nutzen — Perspektiven. *Revista de estudos alemães* 4: 176-199.
- Meyer, C.M. and I. Gurevych.** 2010. Worth its Weight in Gold or yet Another Resource — A Comparative Study of Wiktionary, OpenThesaurus and GermaNet. Gelbukh, A. (Ed.). 2010. *Computational Linguistics and Intelligent Text Processing. Proceedings of the 11th International Conference on Intelligent Text and Computational Linguistics, CICLing 2010, Iasi, Romania, March 21–27*: 38-49. *Lecture Notes in Computer Science* 6008. Berlin/Heidelberg: Springer.
- Meyer, C.M. and I. Gurevych.** 2012. Wiktionary: A New Rival for Expert-built Lexicons? Exploring the Possibilities of Collaborative Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 259-291. Oxford: Oxford University Press.
- Pajzs, J.** 2009. On the Possibility of Creating Multifunctional Lexicographical Databases. Bergenholtz, H., S. Nielsen and S. Tarp (Eds.). 2009. *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*: 327-354. Bern: Peter Lang.
- Pociello, E., E. Agirre and I. Aldezabal.** 2011. Methodology and Construction of the Basque Word-Net. *Language Resources and Evaluation* 45(2): 121-142.
- Resnik, P., M.B. Olsen and M. Diab.** 1999. The Bible as a Parallel Corpus: Annotating the 'Book of 2000 Tongues'. *Computers and the Humanities* 33(1-2): 129-153.
- Reuter, D.** 2010. Interkulturelles Lernen am Beispiel baskischer Muttersprachler im Deutschunterricht. Jarillot, C. (Ed.). 2010. *Bestandsaufnahme der Germanistik in Spanien: Kulturtransfer und methodologische Erneuerung*: 261-266. Bern: Peter Lang.
- Reuter, D. and J. Wolff.** 2007. *Deutsch–Euskaldunentzat*. Donostia: Erein.
- Sanz Villar, Z.** 2013. Hacia la creación de un corpus digitalizado, paralelo, trilingüe (alemán–español–uuskera). Sinner, C. and D. Van Raemdonck (Eds.). 2013. *Fraseología contrastiva del alemán y el español. Traducción y lexicografía*: 43-58. *Études Linguistiques Linguistische Studien* 11. München: Peniope.
- Schmid, H.** 1995. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop, Dublin, Ireland*: 47-50.
- Schmid, H. and F. Laws.** 2008. Estimation of Conditional Probabilities with Decision Trees and an

- Application to Fine-Grained POS Tagging. Scott, D. and H. Uszkoreit (Eds.). 2008. *COLING-08, Proceedings of the 22nd International Conference on Computational Linguistics, 8–22 August 2008, Manchester, UK. Vol. 1: 777-784*. Manchester: COLING.
- Tarp, S.** 1995. Wörterbuchfunktionen: Utopische und realistische Vorschläge für die bilinguale Lexikographie. Wiegand, H.E. (Ed.). 1995. *Studien zur zweisprachigen Lexikographie mit Deutsch II*. Germanistische Linguistik 127-128: 17-61. Hildesheim/New York: Olms.
- Tarp, S.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28(1): 253-267.
- Tarp, S. and H. Bergenholtz.** 2005. Wörterbuchfunktionen. Barz, I., H. Bergenholtz and J. Korhonen (Eds.). 2005. *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*: 11-26. Frankfurt: Peter Lang
- Teubert, W.** 2002. The Role of Parallel Corpora in Translation and Multilingual Lexicography. Altenberg, B. and S. Granger (Eds.). 2002 *Lexis in Contrast: Corpus-Based Approaches*: 189-214. Amsterdam: John Benjamins.
- Wolfer, S., A. Koplénig, P. Meyer and C. Müller-Spitzer.** 2014. Dictionary Users Do Look up Frequent and Socially Relevant Words. Two Log File Analyses. Abel, A., C. Vettori and N. Ralli. (Eds.). 2014. *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, Bolzano/Bozen, Italy, July 15–19, 2014*: 281-290. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism.
- Wolff, J.** 2010. Sprachvergleich Baskisch/Deutsch und seine Auswirkungen für den Unterricht. Jarillot, C. (Ed.). 2010. *Bestandsaufnahme der Germanistik in Spanien: Kulturtransfer und methodologische Erneuerung*: 37-42. Bern: Peter Lang.
- Wolski, W. and A. Cyffka.** 2011. *PONS Großwörterbuch Deutsch als Fremdsprache*. Stuttgart: PONS.
- Zubillaga, N.** 2013. Alemanetik euskaratutako haur- eta gazte-literatura: zuzeneko nahiz zeharkako itzulpenen azterketa corpus baten bidez. PhD Thesis. Vitoria-Gasteiz: UPV-EHU.

Developing a Campus Slang Dictionary for the University of Botswana

Thapelo J. Otlogetswe (*otlogets@mopipi.ub.bw*)
and

Goabilwe Ramaeba (*goabilwe.ramaeba@mopipi.ub.bw*),
Faculty of Humanities, University of Botswana, Gaborone, Botswana

Abstract: This paper focuses on the study of slang on a university campus for a lexicographic project. The research was conducted at the University of Botswana, a campus comprising circa 16,000 students, most of whom are bilingual in Setswana and English, and a small population of foreign students. Very few studies and documentations of slang have been attempted in an African context (cf. Alimi and Arua 2008; Arua and Alimi 2009), in particular for lexicographic purposes. It is therefore hoped that this study will contribute to filling this paucity of literature. In general, university campuses are rich in slang since they consist of a closed group of persons. Slang is the use of informal words and expressions that are not considered standard in the speaker's language or dialect. It is not just informal but it is also considered a lower register of the standard language, therefore disapproved of by users of standard language. In addition to being common within a specific group, in this case a student group, slang is also general within an ingroup such as a gang or a clique. Because of it being proscribed, slang is usually absent from many African language dictionaries which are largely prescriptive in their documentation of the language (Otlogetswe 2006). In this paper, the design of a dictionary of campus slang is proposed that would capture the wealth of the slang of a linguistic community. The data used for the proposed dictionary was collected on the University of Botswana campus.

Keywords: SLANG, SETSWANA, CAMPUS, LEXICOGRAPHY, DICTIONARIES, JARGON

Opsomming: Ontwikkeling van 'n universiteitskampuswoordeboek van sleng aan 'n Afrika-universiteit. Hierdie artikel fokus op die studie van sleng op 'n universiteitskampus vir 'n leksikografiese projek. Die navorsing is gedoen by die Universiteit van Botswana, 'n kampus bestaande uit ongeveer 16,000 studente, die meeste van wie tweetalig in Setswana en Engels is, en 'n klein bevolkingstal vreemde studente. Baie min studies en optekeninge van sleng is in 'n Afrikakonteks onderneem (vgl. Alimi en Arua 2008; Arua en Alimi 2009), veral vir leksikografiese doeleindes. Daar word dus gehoop dat hierdie studie sal bydra tot die aanvul van hierdie gebrek aan literatuur. Oor die algemeen is universiteitskampusse ryk aan sleng omdat hulle uit 'n geslote groep persone bestaan. Sleng is die gebruik van informele woorde en uitdrukkings wat nie as standaard in die spreker se taal of dialek beskou word nie. Dit is nie net informeel nie, maar word ook van 'n laer register as die standaardtaal beskou, en gevolglik afgekeur deur mense wat standaardtaal praat. Benewens dat dit gewoon is in 'n spesifieke groep, in hierdie geval 'n studentegroep, is sleng ook algemeen in 'n binnegroep soos 'n bende of 'n kliek. Omdat dit veroor-

deel word, is sleng gewoonlik afwesig in baie Afrikataalwoordeboeke wat grootliks voorskriftelik in hul dokumentasie van die taal is (Otlogetswe 2006). In hierdie artikel word die ontwerp van 'n woordeboek van kampussleng voorgestel wat die rykdom van die sleng van 'n taalgemeenskap ondervang. Die gegewens gebruik vir die voorgestelde woordeboek is op die kampus van die Universiteit van Botswana versamel.

Sleutelwoorde: SLENG, SETSWANA, KAMPUS, LEKSIKOGRAFIE, WOORDEBOEKE, JARGON

1. What is slang?

Although many have attempted to define slang, it remains a difficult concept. The *Oxford English Dictionary* (OED, Online Edition) describes *slang* as a "language of a highly colloquial type, considered as below the level of standard educated speech, and consisting either of new words or of current words employed in some special sense". Webster (Third Edition) defines slang as "a nonstandard vocabulary composed of words and senses characterized primarily by connotations of extreme informality and usu. a currency not limited to a particular region and composed typically of coinages or arbitrarily changed words, clipped or shortened forms, extravagant, forced, or facetious figures of speech, or verbal novelties usu. experiencing quick popularity and relatively rapid decline into disuse". In the preface to the *Dictionary of American Slang*, Wentworth and Flexner (1960: vi) define slang as "the body of words and expressions frequently used by or intelligible to a rather large portion of the general American public, but not accepted as good, formal usage by the majority". Writers of style manuals especially have generally taken a negative view of slang. Foerster and Steadman (1941: 290) characterise slang as "a cheap substitute for good diction," which demonstrates "laziness in thought and poverty of vocabulary". Fernald (1918: 248) points out that "slang ... saves the trouble ... of thinking. The same cheap word or phrase may be used for any one of a hundred ideas ... Slang is the advertisement of mental poverty."

Slang is therefore a variety of language that is informal and does not conform to convention of vocabulary. Landau (1984: 24) defines slang as "words or expressions that originated in cant (the familiar, non-technical vocabulary restricted to a particular occupation, age, group or any group sharing a special interest). Dumas and Lighter (1978) identify four criteria which are definitive of slang. They argue that an expression should be considered as *true slang* if it meets at least two of these criteria. These are: (1) Its presence will markedly lower, at least for the moment, the dignity of formal or serious speech or writing ... and an individual who has some familiarity with the expression will not expect to find it in the midst of a serious discourse in otherwise standard English except for one special rhetorical effect: to signal that the speaker or writer is deliberately being undignified or intimate with his audience. (2) Its use implies the user's special familiarity either with the referent or with that less statusful

or less responsible class of people who have such special familiarity and use the term. This "special familiarity" usually implies disdain for what is conventionally accepted or esteemed, or an overfamiliarity with what the dominant society finds unseemly or unacceptable. (3) It is a tabooed term in ordinary discourse with persons of higher social status or greater responsibility. (4) It is used in place of the well-known conventional synonym, especially in order (a) to protect the user from the discomfort caused by the conventional item or (b) to protect the user from the discomfort or annoyance of further elaboration. Slang is clearly difficult to define though it is not always difficult to identify.

2. The difference between slang and jargon

There is need however to distinguish *slang* from *jargon* so that in documenting slang, we do not document campus jargon. While a university campus may be rich in slang, it is equally rich in academic jargon which may be mistaken for slang. Southerland and Katamba (1996) define *jargon* as "a label for vocabulary peculiar to some field; occupational sociolect or language of a particular occupational group". A jargon is therefore terminology which is defined in relation to a specific activity, profession, domain, or event. Therefore the following abbreviations, though obscure to persons outside an academic campus, do not constitute slang, but rather university jargon: *C.A.* (continuous assessment mark or term mark from tests and assignments), *SDF* (staff development fellow), *FD* (fail and discontinue). Other unique terms to a university campus include *add and drop* which is an action of adding or dropping courses one wants to do in a semester, *legal clinic* which is a centre for free legal advice that is manned by law students, *e-learning* which is a system whereby students study or communicate with lecturers and others through technological equipment and *freshers' ball* which is a music concert organised by students to welcome new students to the university. All these are part of university jargon and not slang and should not be entered in a campus dictionary of slang. Additionally, they do not qualify as slang, because they are neither colloquial nor lower register.

3. Justification for a dictionary of slang

Lexicography as a discipline is two-pronged. At one level it is the science and art of dictionary making with a general preoccupation of compiling dictionaries. This is known as practical lexicography. At another level it is what is known as metalexicography, that is, the theoretical and critical arm of lexicography, what Svensén (2009) calls "lexicography about lexicography", that considers the theory and criticism of dictionaries and dictionary theory.

Since lexicography deals with the documentation of all varieties of a language and slang constitutes one of the language varieties, slang is therefore an interesting and important area of study for both practical and theoretical lexicographers.

The linguistic and lexicographic argument of this paper is therefore that all linguistic variations, whether formal or informal, must be documented and analyzed to better and fully understand the linguistic variations that exist in a speech community. This is in part because language changes and it may be informative to track such changes over time. It may be informative to determine if there are any patterns of how slang terms have been constructed, that is, whether the terms are drawn from technological developments such as *microfiche* and *drive* or whether the students reach for their cultural peculiarities such as *mopako* and *mokwele* to create slang terminology. Slang documentation is therefore critical since it would document and preserve the different shades of meanings of slang words that exist or have existed and account for how the meanings of certain words have changed over time as well as consider the factors that influenced such changes of meaning.

Slang generally belongs to a closed community and the speakers view it as a part of their identity. Therefore, if a slang word becomes accepted in the "normal" language, the group feels the need to quickly replace it with another unfamiliar one. As a result, slang changes very quickly and that is why it is essential to document the meanings of the words before they are lost.

The documentation of slang will help to show its growth and continuity as a variety of language. It will help to establish if it undergoes the changes that other language varieties go through, such as the coining of new words for new phenomena as a result of technological advancement, new behaviour or attitudes (Aarts and Meijs 1990). English often has new words added to its vocabulary. For example, words such as *Facebook* and *Tweeter* with corresponding nouns *facebooking*, *tweets* and *tweeting* and verbs such as *facebook* and *tweet*. The documentation of slang will therefore help keep track of issues that influence the creation of the slang terminology of a particular group.

Generally it has been observed that the creation of slang words is not confined to any particular aspect of the users' life but that it covers a range of aspects. According to Dumas and Lighter (1978), Hayakawa (1941) called slang "the poetry of everyday life" and argued that it "vividly expresses people's feelings about life and about the things they encounter in life". This seems to be an accurate reflection of what occurs concerning slang amongst the University of Botswana students. Slang terminology covers their academic, sexual as well as economic life. Alimi and Arua (2008) have identified nine semantic categories under which University of Botswana students' slang can be classified. Some of these include: excellence and poorness, absenteeism, grades, cheating, relationships and the competence and incompetence of their lecturers. This paper's interest is to highlight the importance of documenting and preserving slang as a language variety and to suggest how this could be done.

It therefore considers both practical lexicography and metalexicography. Metalexicography is concerned with conceptual areas of lexicography, dictionary design, dictionary criticism, headword selection, the treatment of idioms, proverbs, verbs, nouns, adjectives, etc. in dictionaries. It is also concerned with

developing theories regarding the structure and semantic relationships among words in a dictionary.

4. The user scope of the dictionary

The proposed dictionary is largely written to document the University of Botswana campus slang. Its main target users are University of Botswana students, in particular new students who are unfamiliar to campus life, and therefore may be uninitiated in campus slang usage. It is hoped that it will help aid the users to quickly become part of the student community, giving them a sense of belonging in a new environment. The publication will also be beneficial to university employees who may be unfamiliar with student slang. The main purpose of the dictionary is to give the meanings of words and where possible, to provide the etymology of some of the entries as well as give examples sentences which demonstrate how the words are used in context. It is hoped that the dictionary will also be of interest to sociolinguistics researchers interested in the language of subcultures within a larger linguistic community. University slang would therefore present such important data for linguistic analysis.

5. The proposed dictionary type

The University of Botswana slang consists of several languages, namely Setswana, English and borrowings from other African languages. Because of the bilingualism of the speech community, the dictionary will be unique in that it will comprise words from multiple languages. Inevitably the majority of the words collected are in English, while the rest are mostly in Setswana. The dictionary's entries will therefore be in both English and Setswana, including code-switching and code-mixing, since its target users are largely Setswana speakers who are bilingual in English and Setswana. The definitions of all the entries will be offered in English. The plan is not to make university slang part of a monolingual or a bilingual general dictionary since campus slang is too specialized and not part of the general language.

6. Data and data analysis

Most of modern lexicographic data collection and corpus analysis studies use corpora for collecting data, favouring corpora of naturally occurring language. Biber, Conrad and Reppen (1998: 4) have demonstrated that corpus-based analysis is advantageous, because "it is empirical, it analyses actual patterns of use in natural contexts, it makes extensive use of computers" and is capable of providing quantum information from "large and principled collection of natural texts".

Since slang rarely occurs in writing and is usually characteristic of spoken language, a spoken language corpus would be most appropriate for the

studying of slang on campus. The development of a corpus is however an expensive undertaking and is yet to be developed for the study of University of Botswana slang. Instead, material for this study was collected through a simple questionnaire which asked students to list campus slang terms and their meanings (cf. Arua and Alimi 2009). A total of 1067 respondents were reached through this method. The material was collected, sorted and analysed. It was found that some students confused slang terms with terms which, though colloquial, did not constitute slang. These are colloquialisms such as *babe*, *cool* and *sharp* which cannot be regarded as slang.

Below we present a sample of the collected terms and demonstrate how they could be presented in a dictionary format. In the data presentation the following abbreviations have been utilized: *n.* (a noun part of speech marker), *v.* (a verb part of speech marker) *abbr.* (marks entries which are abbreviations). Idiomatic expressions carry an [*idiom*] label. There are also other stylistic usage markers which are bracketed following a part of speech marker. The sample comprises the following usage stylistic markers: (derogatory), (facetious or derogatory) and (facetious, but with strong derogatory connotations). Illustrative sentences are given in italics to exemplify an entry's usage in context. The data that follows is presented as an illustrative sample of the kind of slang data that was collected at the University of Botswana.

ancestor *n.* (derogatory) 1. an older student who has worked before joining the university for further studies 2. a student who has overstayed their time at the university, especially through retaking and repeating courses. *If you don't want to be an ancestor you better finish your program soon = mature.*

barracks *n.* old white single storey dormitories on campus that are disliked by students. *His room is in the barracks.*

caravan *n.* (facetious or derogatory) a large behind or big buttocks. *John says his girlfriend's caravan is the one thing that attracted him to her.*

caught by a speed trap [*idiom*] to be caught while copying during an exam or a test. *Masego has been caught by a speed trap during our Linguistics test.*

crack *n.* an intelligent student who studies hard. *Thabang is a crack and his friends who do not study hard take advantage of him.*

dog *n.* a young guy; a lad; a chap. *Hi dog! What's up?*

drive *v.* to copy during a test or an examination. *Drive* here refers to a computer drive, therefore one having access to information in a computer drive. *Lebo was caught with a drive in the Chemistry exam.*

drop hole *n.* (derogatory) a loose girl who sleeps around. *Thato is such a drop hole, almost every boy in my residence block has slept with her.*

dubbing *v.* copying from material brought into a test or an examination. *Dubbing*, one of the older terms used on campus, has its roots in the copying of tapes using tape recorders. *Peter has to retake his Physics course because he was caught dubbing during the exam and was disqualified.*

exile *v.* 1. to ask a roommate to find alternative accommodation when your boy-or girl-friend is spending the night over in your shared room. To make

another student leave the room they share and spend the night elsewhere is here compared to one being exiled from their own land. *Tshepo exiles his roommate every weekend.*

fax v. to slip an assignment under a lecturer's door in their absence. This is usually done in the evening in the absence of the lecturer by students who submit their assignments late and want to avoid facing the lecturer. *Dr. Tshane hates it when students fax her their assignments.*

fax out of order [idiom] this expression is used when there is no room under the door through which an assignment could be slipped into a lecturer's office. *The students were unable to fax their assignments to Dr Tshane because the fax was out of order.*

fish market n. (facetious, but with strong derogatory connotations) female residences. *On Friday nights the fish market is always buzzing with young men hoping for a good catch.*

fish pond n. (facetious, but with strong derogatory connotations) female residences. *On Friday nights the fish ponds are always buzzing with young men hoping for a good catch.*

go ja mophato [idiom] to repeat. The word *mophato* is a noun that refers to an age group, traditionally who join the initiation school together. The expression therefore literally means to drop out of a class of your age group. *John o jele mophato too many times and eventually he quit university.*

go thuba mmopa [idiom] to talk nonsense or to fail a course completely. The phrase literally means "to break clay". *Mpho had to drop his Psychology course because o ne a thuba mmopa.*

go tshela class metsi [idiom] to dodge or to miss a class. The expression liter-

ally means "to pour water on a class". *I had to finish and submit my Sociology assignment, so I had to tshela metsi on my other classes to finish it on time.*

gold mine n. an older guy showering young women with gifts and money. *Lulu's gold mine bought her an expensive phone on Valentine's day.*

high risk n. a married man or woman in a sexual relationship with a student. *Students like high risks because they shower them with expensive gifts and money.*

jesa v. (derogatory) to lecture poorly. The word means "to feed", so here it is used ironically to show that the lecturer is not really "feeding" the students anything worthwhile. *Students do not like junior lecturers, they say ba a jesa.*

July conference n. a supplementary examination held in July at the end of the academic year. *Many students from the Faculty of Science came for the July Conference but many of them still failed their courses.*

kgaola v. to go: *A re kgaole or we'll be too late for class.* The word literally means "to cut" in Setswana.

koko ya Setswana n. (derogatory) a rural girl or boy who is resistant to modernisation: *She is too much of a koko ya Setswana; she will not change her hairstyle and dress sense no matter how hard you try.* The phrase means "a free-range local Setswana chicken".

lamza v. to kiss: *I wish I could lamza that girl, she is very beautiful.* The word is not Tswana, but a student creation.

Las Vegas n. one of the best looking accommodation blocks on campus considered by students to be modern and of a high standard. *Every student on campus wished they stayed in Las Vegas.*

legosha *n.* (derogatory) a female student who uses sex to manipulate men to get money from them. *Neo is a legosha, she sleeps with men for money.*

leshalala *n.* a girl who maintains high standards of living: *That girl ke leshalala, she would do anything to keep up her high standard of living.*

le-year year *n.* a third year student and above. It refers to one who has spent many years as a student. *They found le-year year intimidating first year students at the bookstore.*

madoda score *n.* the 50% mark in a test or an examination. The Ndebele word *madoda* means "men" or "gentlemen". *John is a hopeless student, he passes all his courses with a madoda score.*

majesa *n.* (derogatory) a lecturer who lectures poorly: *Our Maths lecturer is such a majesa.*

many more *n.* blue movies: *They've been watching many more the whole weekend.*

matratane *n.* a skinny person. (from the Setswana word 'terata' which means a wire. *That girl is such a matratane, you would think she was sickly.*

mature *n.* (derogatory) 1. an older student who has worked before joining the university for further studies. *Most mature students work very hard and pass their courses well.* 2. a student who has overstayed their time at the university, especially through retaking and repeating courses. *If you don't want to be a mature you better finish your program soon = ancestor.*

mchine *n.* a corruption of *machine*; anything of quality; also, an intelligent person. *David is a mchine, he passed all his courses with distinction.*

microfiche *n.* a page with notes which has been folded multiple times until it is tiny, sneaked into a test or an examination room. *Sam was disqualified*

from writing the exam because he was caught with a microfiche.

mmamochachose *n.* a lady who sells food in the street just outside the university campus. The name comes from a chicken outlet known as *Mochachos*. *Most students prefer to buy lunch from mmamochachose because refectory food is expensive.*

mokwele *n.* a male student who is afraid or too shy to approach girls. The word means "a sheep that is given as part of the bride price". *Tshepo is a mokwele, he does not have a girlfriend.*

mopako *n.* material brought into the test or examination by a student from which they copy: *Many students failed the exam, because they were caught with mopaku.* The word means "packed food that are taken to school, to work or on a journey".

OA [*abbr.*] (derogatory) old age, used to refer to older students: *He is an OA, but he is only studying for his first degree now.*

others *n.* students who are not doing science-based courses or Law. *Students who study science courses look down upon others because they think that they are not intelligent.*

phoko *n.* a young male who is considered a hero by others because of what he has done or what he does (such as sexual prowess or academic strength). The word in Setswana means "a billy goat". *James is a phoko, he has slept with most of the girls in my residence block.*

roomza *n.* room mate. *Sue and her roomza Pet are best friends.*

sack man *n.* a male student without a girl-friend. *John is a hopeless sack man who cannot find a girlfriend.*

sethubi *n.* a studious student. The word in Setswana means "one who breaks",

implying that such a student breaks down complex ideas. *Lala is a sethubi, she passed all her exams with flying colours.*

sexually transmitted degree *n.* degree acquired as a result of having a sexual relationship with a lecturer. *Neo had sexual relationships with most of her lecturers, so hers is a sexually transmitted degree and not a genuine one.*

sexually transmitted marks *n.* marks acquired as a result of having a sexual relationship with a lecturer. *Neo is Prof Dambe's girlfriend, so she passes because of sexually transmitted marks.*

single *n.* a room in which one stays alone, allocated only to completing students. *Only fourth year students can stay in singles.*

squatter boy or girl *n.* an off-campus student who sleeps on campus with another student in their room without the approval of university officials. *Many students want to stay off-campus to get the monthly allowance, but then they inconvenience other students by becoming squatter boys and girls in their rooms.*

SRB [*abbr.*] (facetious or derogatory) strong rural background. *I like that girl, but she has a high SRB that turns me off. My brother has too much of a SRB, he wouldn't marry outside his race.*

STD [*abbr.*] sexually transmitted degree. *Neo had sexual relationships with most of her lecturers so hers is a sexually transmitted degree and not a genuine one.*

swipe *v.* to have casual sex with someone not your boy- or girl-friend. *The youth like to swipe a lot, despite the high rate of HIV/AIDS in the country.*

thuba *v.* to study very hard. The word literally means "to break". *Tshepo wa thuba, although he also parties hard.*

tse di fofang *n.* very beautiful girls. The expression means "those that fly", probably in reference to birds. *Most of the girls who stay in my block ke tse di fofang.*

veteran *n.* (derogatory) one who has overstayed their time at university, because of failing and repeating courses. *Mike is such a veteran, he has been at this university for a very long time.*

write dry [*idiom*] to write an examination or a test without having prepared for it. *Peter went partying last night instead of studying for his PE test so he wrote it dry this morning.*

year year *n.* a student who is doing third, fourth or fifth year: *Years years like intimidating new students.*

zero-grazing *adj.* Referring to a student who lives on campus, but does not take their meals from the refectory. This in particular applies to students who are receiving government sponsorship. It is perceived by students as derogatory since a zero-grazing student is one who is stingy and not wishing to spend money on food, preferring to use it for other matters. *Zero grazing students are always tired and sleepy because they do not eat well.*

The data presented above covers some of the general themes that are typical of a campus environment such as sexual relations between students and lecturers as well as perceptions associated with sexuality on campus. For instance the expression *a drop hole* refers to a female student of lower morals, *to exile* another student is to make them spend the night away from their shared room giving space to their roommate to enjoy privacy with a partner. *A sack man* is a young

man lacking a girlfriend while a *legosha* is a female student who exchanges sexual favours for money from men. *Sexually transmitted marks* and *sexually transmitted degrees* point to the unethical conduct of students extending sexual favours to lecturers in exchange for favourable grades.

The data also comprises words which expose academic dishonesty by students. For instance *mopako* is material brought into a test or examination by a student from which they copy while the act of copying is termed *dubbing*, clearly a term which has its roots in the tape recording period while *drive*, which also means copying, has its roots in a computer drive. The idiom *caught by a speed trap* articulates being caught copying in a test or examination.

Other terms articulate the presence or lack of academic brilliance. *Go thuba* literally meaning to break in Setswana, however in campus slang it means to study hard. An act of studying in this case is perceived as breaking down complex university material into something easy to digest. *Jesa*, literally meaning feed, is used to characterise a lecturer who delivers his or her material poorly. Such a lecturer in slang terminology is known as *majesa* — one who feeds. *Go ja mophato* is an idiom which is used to mean to repeat a year. The idiom is derived from Setswana culture. The literal meaning of the idiom is 'to avoid going to initiation school'. It is therefore evident that going to school is compared to going to initiation school. One who is academically brilliant is termed a *crack*, a term which is related to *go thuba* above.

From the data presented, there is evidence that slang demonstrates high levels of creativity. It attaches new meanings to commonly established words that are in general use. For instance, the word *mature* in general use has positive connotations of one who does things following careful consideration; however, in university slang, it has negative connotations since it means "a student who has overstayed their time at university, especially through retaking and repeating courses". The word *fax*, a short form for *facsimile*, is the transmission of scanned text or images to a telephone number connected to a printer or other devices. In university slang, it now negatively means to slide paper material such as an assignment under the door of a lecturer in their absence. Creativity is also visible in the use of fairly well established abbreviations such as STD which in normal usage stands for *sexually transmitted disease* but in the University of Botswana slang stands for *sexually transmitted degrees*, something that is perceived as an ill-gotten degree through an exchange of sexual favours with staff for good course marks.

Lexicographically, lexicographers are interested not just in new forms or new words; they are also interested in new meanings associated with established words. Such meanings may be caused by dialectal variation or stylistic differences. Slang creativity is also demonstrated by code-mixing in which a single word contains parts of two or more languages; in our case, Setswana and English. For instance, the word *le-year year* has a Setswana prefix *le-* usually associated with nouns of class five of the Setswana noun class system, so that *year year* then refers to 'many many years'. In other instances, the same meaning

associated with a common word is then reapplied in a new context. For instance, *dubbing* means copying tape contents into another tape using a tape recorder. However in slang, it is used to mean copying from illegally brought in material in a test or examination situation. There are also idiomatic expressions such as *go thuba mmopa* "lit: to break mud" and *be caught by a speed trap* which are clearly semantically non-compositional and absent in both common expressions of Setswana and English respectively.

From our data analysis, it has become clear that slang terms and expressions used at the University of Botswana are fairly restricted since they are mainly about campus matters such as studying, copying during a test, sexual relationships between students and staff and a few matters relating to campus life. Lexicographically there is a need to deal with dictionary compilation of slang carefully. Decisions must be taken whether the same dictionary will have all the slang headwords though they are from different languages. Will such a dictionary be conceived as a monolingual or bilingual dictionary? Additionally, in the design of a slang dictionary, though slang is informal, lexicographers would have to consider stylistic markers such as *offensive* or *archaic* to label terms which fit such labels.

The target users of the proposed dictionary is therefore the University of Botswana students, especially new and foreign students who usually find themselves lost in the web of university slang. Creating a slang dictionary will also document slang usage over the years and provide useful linguistic data of how developments in technology and social life have influenced the development of slang on campus.

7. Conclusion

The paper has argued for the importance of the documentation of slang as a language variety of a specific language community. The data has revealed that slang at the University of Botswana exists in both English and Setswana and sometimes the two languages are combined to form a slang expression. Slang like other language varieties is not stagnant; it grows and new terms are frequently added to its lexicon which justifies the need for its documentation. For instance, the expressions *microfiche* and *many more* were common in the 1990s and have largely fallen into disuse. The recent expressions from the year 2005 include amongst others *zero-grazing*, *tse di fofang*, *caught by a speed trap* and *Las Vegas*. Slang is as much a language variety as any other language variety and therefore deserves to be documented and studied descriptively. Since slang generally does not last long, particularly on a campus where students are annually arriving and leaving, there is a need to devise systems and strategies to document it. This paper has argued that the practice of documenting slang is attractive since it will record the shades of meaning over a period of time, changes in word usage and new words that have been created.

References

- Aarts, J. and W. Meijs (Eds.).** 1990. *Theory and Practice in Corpus Linguistics*. Amsterdam: Rodopi.
- Alimi, M.M. and A.E. Arua.** 2008. Gender and Student Slang in the University of Botswana. Arua, A.E., M.M. Bagwasi, T. Sebina and B. Seboni (Eds.). 2008. *English Language and Literature. Cross-Cultural Currents*: 38-53. Cambridge: Cambridge Scholars Publishing.
- Arua, A.E. and M.M. Alimi.** 2009. The Creation of Students' Academic Slang Expressions in the University of Botswana. *Linguistik Online* 40(4).
- Biber, D., S. Conrad and R. Reppen.** 1998. *Corpus Linguistics: Investigating Language Structure and Usage*. Cambridge: Cambridge University Press.
- Dumas, B.K. and J. Lighter.** 1978. Is Slang a Word for Linguists? *American Speech* 53(1): 5-17.
- Fernald, J.C.** 1918². *Expressive English*. New York: Funk & Wagnalls.
- Foerster, N. and J.M. Steadman.** 1941. *Writing and Thinking: A Handbook of Composition and Revision*. Revised edition. Boston: Houghton Mifflin.
- Hayakawa, S.I.** 1941. *Language in Action*. New York: Harcourt, Brace.
- Landau, S.I.** 1984. *Dictionaries: The Art and Craft of Lexicography*. Cambridge/New York: Cambridge University Press.
- Otlogetswe, T.J.** 2006. Challenges to Issues of Balance and Representativeness in African Lexicography. *Lexikos* 16: 145-160.
- Oxford English Dictionary (OED, Online Edition)*. <http://www.oed.com/view/Entry/181318?rskey=WcfYK5&result=3&isAdvanced=false#eid> [Accessed 16 January 2013].
- Southerland, R.H. and F. Katamba.** 1996. Language in Social Contexts. O'Grady, W., M. Dobrovolsky and F. Katamba. 1996. *Contemporary Linguistics: An Introduction*: 540-590. London/New York: Longman.
- Svensén, B.** 2009. *Practical Lexicography. Principles and Methods of Dictionary-Making*. Oxford/New York: Oxford University Press.
- Webster's Third New International Dictionary of the English Language Unabridged*. 1961. London: G. Bell and Sons/Springfield, Mass.: G. and C. Merriam.
- Wentworth, H.B. and S.B. Flexner.** 1960. *Dictionary of American Slang*. New York: Thomas Y. Crowell.

Key Issues in Fuertes-Olivera and Tarp's *Theory and Practice of Specialised Online Dictionaries*

Valeria Caruso, *Department of Literary, Linguistic and Comparative Studies, University of Naples 'L'Orientale', Naples, Italy*
(vcaruso@unior.it)

Abstract: Fuertes-Olivera and Tarp's *Theory and Practice of Specialised Online Dictionaries* is a long-awaited volume, offering a systematic description of the tenets of function theory. The authors present this theoretical framework both as a guideline in order to accomplish different lexicographical tasks (i.e. from compiling dictionaries to assessing the lexicographical suitability of other theoretical models), and as a valuable methodological tool which can be used to detect the users' needs. The 'Functional approach' is thus considered to be an alternative method to the research on users' behaviour. However, not only the new ideas and the terminology used within this field of lexicographical research deserve attention, but also some issues related to the concept of specialised lexicography, which are briefly outlined by Fuertes-Olivera and Tarp in the book. Their definition of this sub-discipline stimulates further reflections on the epistemic status of the two branches of lexicography (general and specialised) as well as on the related concepts of general and cultural knowledge.

Keywords: DICTIONARY CRITICISM, ENCYCLOPAEDIAS, SPECIALISED KNOWLEDGE, SPECIALISED LEXICOGRAPHY, LEXICOGRAPHICAL THEORY, ONLINE DICTIONARIES, PROSCRIPTION

Opsomming: Sleutelkwessies in Fuertes-Olivera en Tarp se *Theory and Practice of Specialised Online Dictionaries*. Daar is lank gewag vir Fuertes-Olivera en Tarp se *Theory and Practice of Specialised Online Dictionaries* — 'n boek wat 'n sistematiese beskrywing bied van die beginsels van die funksieteorie. Die skrywers bied hierdie teoretiese raamwerk sowel as 'n riglyn om verskillende leksikografiese take uit te voer (van die samestelling van woordeboeke tot by die beoordeling van die leksikografiese gepastheid van ander teoretiese modelle) as 'n waardevolle metodologiese instrument wat gebruik kan word om gebruikersbehoefes vas te stel. Die funksionele benadering word dus beskou as 'n alternatiewe metode tot navorsing oor gebruikersgedrag. Dit is egter nie net die nuwe idees en terminologie wat in hierdie veld van die leksikografie gebruik word wat aandag verdien nie maar ook sekere kwessies verwant aan die begrip van vakleksikografie wat kortliks deur Fuertes-Olivera en Tarp in hierdie boek aangebied word. Hulle definisie van hierdie subdissipline stimuleer verdere nadenke oor die epistemiese status van die twee vertakings van die leksikografie (algemene en gespesialiseerde) asook oor die verwante begrippe van algemene en kulturele kennis.

Sleutelwoorde: AANLYN WOORDEBOEKE, ENSIKLOPEDIË, GESPELISEERDE KEN-

NIS, LEKSIKOGRAFIETEOORIE, PROSKRIPSIE, VAKLEKSIKOGRAFIE, WOORDEBOEKKRITIEK

1. Tools and Functions to overcome division

Theory and Practice of Specialised Online Dictionaries, published this year by De Gruyter, is presented by the authors, Pedro A. Fuertes-Olivera and Sven Tarp, as a compendium of the lexicographical research carried out within the tenets of the function theory. Reminiscent of its antecedent, the *Manual of Specialised Lexicography*, the authors underline how 20 years of advances in the field have necessitated an up-to-date report on the theoretical assumptions that in 1995 were still in their infancy. However, not only the theory, but also the actual activity of compiling dictionaries has increased enormously ever since, in terms of both quantity (with more than one hundred vocabularies published) and quality. In this regard, the collaboration between the Centres for Lexicography in Aarhus (Denmark) and in Valladolid (Spain) is recognised as one of the most significant contributions to have enhanced the development of lexicographical theory and practice, and two representatives of these institutions have consequently written this comprehensive volume about the function theory of lexicography.

The book opens (Chapter 1: *Introduction*, and Chapter 2: *What is Specialised Lexicography?*) with historical notes explaining the origins of what the authors consider to be the first "schism" in lexicography: the strict division between dictionaries describing general language, and those explaining "things", namely specialised dictionaries and encyclopaedias. According to Fuertes-Olivera and Tarp, Samuel Johnson is responsible for this division. The British lexicographer seems to have deliberately excluded from the preface to the *Dictionary of the English Language* any assertion regarding the fact that dictionaries could also deal with 'things', on the contrary, he remarked the fact that they are 'word-books'. However, in the 18th century other authors, such as John Harris and Jean le Rond D'Alembert, affirmed that dictionaries could portray terms and sciences as well. Therefore, the dictionary was considered to be an organizing format suitable for whatever kind of knowledge it was required to convey. In fact, this same idea is one of the main assumptions of the lexicographical function theory, as is stated more clearly elsewhere by Tarp:

[t]he truly unique thing about dictionaries is not the various types of data they employ in covering the information needs of users [...]. Such data can generally be incorporated into other types of book and text as well. The truly unique thing is the way in which this data is made accessible so users can quickly and easily find the exact data they need. (Tarp 2008: 101)

The way dictionaries allow users to get quickly to the required data is, thus, the true focus of lexicography, regardless of the nature of the information itself. Data access and a broader reference science ('accessology') were thus (Tarp 2008) advocated as the necessary step forward to update lexicography accord-

ing to the requirements of the new information society. In fact, the broad concept of 'accessology' has inspired much research and many papers in the field of function theory (Bergenholtz and Gouws 2010; Leroyer 2011; Fuertes-Olivera and Niño-Amo 2011, to quote but a few), but it has also caused scepticism outside the realm of this paradigm of studies. Lew (2008), for example, maintained that lexicography should preserve its purview, and should not morph into a wider reference science.

Theory and Practice of Specialised Online Dictionaries approaches the matter from a different angle, with reference to the same topic of broadening the field of lexicography in order to compile dictionaries that, by taking advantage of the new information technologies available, would allow users to satisfy their information needs more easily. Fuertes-Olivera and Tarp choose in fact to discuss critically two epistemic distinctions which have traditionally characterised lexicography. The first is the above mentioned opposition between general and specialised dictionaries, the other is a sub-field discernment regarding the discriminative features of encyclopaedias and specialised dictionaries. Of the two, only the second proves to be inconsistent, while the other is eventually upheld by the authors.

The distinction between general and specialised dictionaries is in fact taken as a starting point to dismantle the confusing labels applied to different types of dictionaries on the basis of their specialisations. This descriptive approach produces an incongruous metalexicographical terminology, since there are too many labels used for naming specialised dictionaries, which are nonetheless still insufficient to cover the huge variety of available resources, i.e. *specialised dictionaries*, *LSP dictionaries*, *technical dictionaries*, *terminological dictionaries* or, lastly, *special dictionaries*. On the contrary, Fuertes-Olivera and Tarp apply their usual method of abstraction to multi-faceted real objects and decide to adopt the traditional distinction between 'general' and 'specialised' lexicography, which is based on the kind of knowledge dealt with in dictionaries (i.e. general or specialised). The choice proves to be advantageous from a practical point of view, explaining for example why dictionaries of collocations (or with other linguistically-restricted coverage) should not be called 'specialised'. At the same time, this juxtaposition is used to formulate the following statement:

specialised lexicography is here defined as the branch of lexicography concerned with the theory and practice of *specialised dictionaries*, i.e. dictionaries, encyclopaedias, lexica, glossaries, vocabularies, and other information tools covering areas outside general cultural knowledge and the corresponding Language for General Purposes (LGP). (7)

The distinction between "general cultural" and "specialised knowledge" is presented as intuitive and uncontroversial, thus it is not elaborated on. The only additional notation provided regards the fact that their opposition is purely transitional, since it is impossible to grasp precise divisions between the realm of specialised and general knowledge. Therefore, no clear-cut divisions can be

outlined between these two blurred categories, which overlap in a grey area. This explanation partially builds a bridge between specialised and language dictionaries, a point that is subsequently clarified in the book, with a specific call for lexicographical theory to focus on the common features of dictionaries, rather than emphasising their bewildering differences.¹

However, it must be admitted that a definition of the kind "X is what Y is not" ("tools covering areas outside general cultural knowledge and the corresponding Language for General Purposes") is rather inelegant and, since the categories involved are transitional, a set of relevant features would have probably been more suited to outline the unstable boundaries of specialized lexicography. On the other hand, it must be acknowledged that 'general knowledge', as well as Language for General Purpose are rather indefinite concepts, which derive from their counterparts, i.e. 'specialized knowledge' and Language for Special Purposes. The opposite is true of the definition provided by Fuertes-Olivera and Tarp, which derives specialised knowledge from the most unclear concept in the discussion, i.e. general knowledge.

This last notion is surely worthy of further analysis, together with the role assigned by the function theory to linguistic facts in its general framework, and with respect to the general lexicography (as opposed to the specialised lexicography). Readers, like myself, might also be puzzled by the role of language provided in the epistemic division, which distinguishes between general and specialised languages, paralleling general and specialised knowledge.

2. Language and lexicographical descriptions within the framework of function theory

From a careful reading of this book and others (Tarp 2008), however, some relevant explanations can probably be collected. Language, in fact, proves to encompass two different ontological statuses in function theory. On the one hand, it is a skill (52-54 and Tarp 2008: 131-136), allowing native speakers "to communicate with other speakers of the same language". On the other, it is a specific kind of knowledge, namely a "*learned knowledge of the language*" consisting of a "conscious, systematic knowledge" of one language, which allows people to explain "the precise meaning of individual words or to formulate the grammatical rules which they are unconsciously automatically able to use" (Tarp 2008: 133). In *Lexicography in the Borderland between Knowledge and Non-knowledge*, Tarp clearly explained that "in a living language vocabulary and grammar do not have their own independent existences" (Tarp 2008: 135), but Linguistics can help a lot since, using abstraction, this science is able to separate and examine each of them, and the

knowledge resulting from this examination can be communicated in connection with the study of a foreign language. This communication can take place by teaching, by the autonomous study of textbooks and grammar books, and by consulting dictionaries. (Tarp 2008: 135)

This explanation offers clear evidence regarding the fact that the Language for General Purpose is a set of metalinguistic generalizations (named by Tarp 2008 'language knowledge') about the grammar and vocabulary used by people belonging to different professional groups in everyday life and interaction, to convey general cultural concepts. The features of this knowledge-oriented description of one language are dealt with in general dictionaries, and they should be identified, as the previous quotations suggest, on the basis of abstraction, using pertinent theories (mainly belonging to the branch of Sociolinguistics, one assumes). It is, therefore, not out of place to cite here one particular area of Linguistics research, carried out in Italy by Tullio de Mauro in the 60s and currently undergoing revision in order to be updated (De Mauro and Chiari 2014).

The aim of this investigation (De Mauro 2004) was to portray the 'basic vocabulary' of the Italian language, namely the minimal inventory of words that are necessary to cover the basic understandings of ordinary life. The research was carried out in order to promote the simplification of bureaucratic language and official texts used in public communication. Results were thus published in the form of a prescriptive wordlist, recommended to be used when compiling official documents, which should be able to be understood by everyone. The project was therefore intended to identify the bulk of the Language for General Purpose of a linguistic community, representing its general cultural knowledge as well. The methodology employed for this project is therefore interesting and relevant also for our discussion, highlighting the kind of problems posed by this type of language delimitation, as well as the procedures adopted to overcome them. Firstly, the concept of 'basic vocabulary' was quantified in terms of the amount of words that are mastered by native speakers with a primary school qualification, i.e. 8 years in the educational system of the country. This led to a selection of approximately 7200 words, the majority of which were derived from the list of the 5 000 highest-frequency words of a corpus, reduced to 4 750 after a comprehension test administered to Italian speakers with a primary school qualification, both adults and children. The remaining 2 300 items are instead 'highly available' words, or words that are rarely used, like the names of common tools (e.g. wipers or mops), albeit they belong to everyday life and thus are part of our general cultural knowledge. They were not extracted by corpora, but were chosen from dictionaries and interviewing different groups of people. Despite the fact that no dictionary would be really useful if it contained only the lemmas that everyone already knows ('basic vocabulary'), the study shows some methods in Sociolinguistics research integrating various approaches (from corpora to interviews and tests) in order to overcome some issues related to the concept of general language and common knowledge. Lastly, the 'basic vocabulary' of one language represents a useful repository of words for writing definitions that everyone can understand.

Going back to the role that linguistic aspects play in lexicographical descriptions according to function theory, it may be inferred from the previous

discussion that language is the topic of lexicographical description as long as it is useful to support the intended users with the tasks for which the dictionary is compiled. In fact, language *per sé* is not at the forefront of the discussion, but its content-related counterparts, i.e. the Language for General or for Special Purposes. For example, collocations and grammar notes should be offered in specialised dictionaries when support with communicative tasks is required, as is shown in Chapter 9 of the book. Likewise, different levels of linguistic descriptions require appropriate theoretical solutions and, thus, the proper metalinguistic analysis, in order to comply with the intended aim of the dictionary, e.g. to describe the history of one language, its contemporary use and forms, or learning one language for native or non-native speakers (26).

3. Typological and evaluative classifications of dictionaries

As the previous discussion highlights, the content-dependent interpretation of linguistic facts (general and specialised language) also dismantles the other troublesome dispute in lexicography presented in the first pages of the book, namely the "Chinese wall" between LSP dictionaries and encyclopaedias. In fact, the authors maintain that specialised lexicography "by far transcends a mere description of the various specialised languages and also treats the very substance of these disciplines themselves in order to provide direct, punctual access to their cognitive achievements" (8). Therefore, the "Chinese wall" traditionally raised between more linguistically-oriented tools (commonly called 'LSP dictionaries') and those relying more heavily on the description of "things" (generally referred to as 'encyclopaedias') is actually inconsistent, provided that these reference works differ only in the type of support offered to users. While encyclopaedias have traditionally dealt with specific topics more extensively, in order to provide a "profound study of one or more subject fields" (12), dictionaries tend to offer more punctual information to assist with different, but more circumscribed tasks: e.g. translation or comprehension of one specific term, spellchecking, choice of the correct collocation. The authors underline that these various activities require different lexicographical solutions in order to be carried out, albeit the intrinsic nature of these tools remains the same. In addition to these arguments, based on the juxtaposition with other studies, the authors offer an illustrative metaphor taken from Diderot's (1755) preface to the French *Encyclopédie*, which illustrates the essence of specialised reference works that Fuentes-Olivera and Tarp have in mind. As the French Illuminist explains, the pioneering work of the *Encyclopédie* was meant to assemble "knowledge scattered over the face of the earth", making it available to all humankind for the purpose of improving education and, consequently, increasing happiness. The ethical standpoint offered by the illuminists, as well as the similarities with their visions and intents honour the lexicographical function theory with new arguments, shifting the focus on different aspects of the core idea of a broader reference science (elsewhere called 'accessology'), whose principles should

inform its subordinated sub-disciplines. However, it must be underlined that this point is not at the forefront of discussion in Chapter 2, since it follows the long presentation of the many epistemic boundaries that prevent a radical shift in the field of lexicography. In this respect, *Theory and Practice of Specialised Online Dictionaries* adopts a more moderate view, avoiding a discussion of the more general and "provocative ideas" (Fuertes-Olivera and Bergenholtz 2011: vii) of function theory, such as the above-mentioned topic of "accessology", which receives an indirect mention only by quoting Henriksen's (1992) "referensology" (8).

On the contrary, many other topics of function theory are re-proposed in the volume, and some have also been enriched by additions and new details. This is the case, for example, with the evaluative typology of electronic dictionaries, which elaborates on some earlier proposals by Tarp (2011, 2012), adding one grade in the evaluation scale, i.e. "Stray Bullets" dictionaries.

These tools, in fact, can be divided into five groups, depending on the kind of support they offer to users, with respect to the technology employed. The assessments range from very poor resources, which are online reproductions of printed editions by means of scans or photographs (Copycats), to the adaptive (Gamper and Knapp 2002) tools of tomorrow, called "Rolls Royces", which should be able to customize themselves according to the user's needs. At the lower end of the scale, there are the "Faster Horses", tools improving paper dictionaries because some search facilities have been added, and the "Stray Bullets", which misuse technology since they do not know exactly what to do with it. In fact, some of these dictionaries make information drop behind the fancy use of technology, transforming the dictionary in an entertaining device (called "lexicotainment" by Almind, Bergenholtz and Vrang 2006). Some others, however, customize their entries in terms of the amount of data displayed, instead of tailoring the kind of information provided on the basis of the user's needs. Both these tools are far from achieving any of the qualitative improvements that are expected to take place in electronic lexicography, and have already begun to become reality, in the view of the authors, with "Model T-Ford" dictionaries, a label named after the revolutionary car produced by Henry Ford. These resources actually allow users to set the entries according to their expertise level, and the specific task to be fulfilled. Among the first tools provided with these functionalities, one can list the "monofunctional" dictionaries of the *Diccionarios de Contabilidad*, edited by the International Centre for Lexicography at the University of Valladolid, and the Centre for Lexicography at Aarhus University.

4. Theoretical status of lexicography and some academic partitions

Chapter 3 (*Academic Status of Specialised Lexicography*) and 4 (*Concept of Lexicographical Theory*) explain why lexicography is an independent science with its own theories.

The discussion moves from evidence regarding the fact that, despite the huge variety of reference works produced over the centuries, there is an intuitive understanding that all of them are lexicographical products. This highlights the fact that some common features should exist, and they are independent from the specific data contained in every single dictionary, thus proving that contents (or data) and containers² (dictionaries or whatever other organizing tool) conduct separate lives. Having considered these matters, lexicography should be independent from any other discipline, in order to promote its own theories and principles for a better production of the ordered containers of knowledge that are commonly called dictionaries. Further proof is also provided to sustain this independence.

First of all, the authors affirm that no reflections inspired by theories in Linguistics have proved to be of any assistance to the solution of real lexicographical problems. On the contrary, experts in one subject field, who were unacquainted with Linguistics, have in fact produced many excellent specialised dictionaries, such as Postlethwayt's and Savary des Bruslons' Economics dictionaries in the 18th century. Linguists have instead collaborated on lexicographical projects that are basically "conceived with a view to documenting a language" (26), however with different aims such as "assisting text reception, production and translation", or "supporting the acquisition of a first and second language". Fuentes-Olivera and Tarp admit that there are "excellent lexicographical works" among these, "yet all of them can undoubtedly be improved" (26). What the authors seem to complain about is the innate blindness of linguists, who are unable to discern between the topic of their lexicographical descriptions (i.e. the language) and the hosting tool (i.e. the dictionary), which is intrinsically different from the other, since it serves other purposes (the best presentation of data to the user). In addition, academic debates are also rendered even more complicated by the presence of terminologists, who claim to be the authorities in the description of terms using terminography. Therefore, those defending a lexicographical theory put into question the balance of stable partitions: linguists compile general language dictionaries, and terminographers edit terminology resources. These divisions prevent from a more congruous combination of disciplines that, according to Fuentes-Olivera and Tarp, should merge Information Science and Lexicography on the basis of their common major concern: how to retrieve information. Fruitful collaborations have already begun within the tenets of function theory, concerning "needs-adapted data presentation in e-information tools" (32), and the reconsideration of a key topic in Information Science like 'relevance', whose concept has been put into question by the lexicographical theory (Bothma and Tarp 2012).

An extended discussion of the theoretical status of lexicography is presented in Chapter 4. Within these pages, the authors demonstrate why lexicography is a science, contrary to the claims of other lexicographers and metalexicographers, who have expressed disbelief or maintained their positions with more detailed statements. The essential points in the debate focus on the fact

that lexicography is a practical activity dealing with the creation of utility tools. Bejoint (2010) has thus questioned, "How can there be a theory of the production of artefacts?", since a theory is assumed to describe natural phenomena. Others, like Bogaards, have asked instead for a falsifiable method, i.e. the fact that every hypothesis ought to be verified in order to be scientific. Fuertes-Olivera and Tarp argue that both positions point to the paradigms of natural (in the case of Bejoint) and hard sciences (with respect to Bogaards), while it is undeniable that many scientific and academic disciplines belong to social phenomena which comply with different laws and principles, and these sciences are generally referred to as 'social sciences'.

Moreover, a theory is needed as long as the practical activity, i.e. compiling dictionaries, needs to be qualitatively improved, something which is possible only by means of sound guiding principles. With respect to these assumptions, the authors derive their epistemological model from the philosophy of science, which conceives theory as an "organised set of statements about an area of objective reality" (37), aiming to demonstrate the similarities among different objects as well as their relationships. Thus, theory is mainly an exercise in abstraction, which can nevertheless be applied at different levels like, for example, the "whole discipline" (*general theories*) or only "a restricted subarea" (*specific theories*). These generalizations may also be part of a general theory (*integrated theories*) or not (*non-integrated theories*); they may finally be purely descriptive of existing practices (*contemplative theories*) or, contrarily, they may be the guidelines for future developments (*transformative theories*; 38). The lexicographical function theory, making abstractions from the millenary dictionary practices, has eventually formulated its theoretical statements, which portray the essence of the lexicographical activity and thus are able to display the guidelines to improve dictionaries.

5. Key concepts in function theory

The well-known statement defining a 'lexicographical function' is offered again and again in the book using different formulations, whilst it is framed in its usual terms in Chapter 5 (*General Theory of Specialised Dictionaries*). Here the authors explain that a "lexicographical function", or the function that every dictionary should fulfil, consists in "the satisfaction of the specific types of punctual information need that may arise in a specific type of potential user in a specific type of extra-lexicographical situation" (64). This chapter is the most relevant in the book, presenting the main tenets of the theoretical paradigm, and stating more explicitly some of its most innovative, and unconventional methodologies. Basically, lexicographical functions are presented as a method used to detect users' needs, and the "Functional approach" is listed among other possible approaches to this topic: "Business as usual", "Personal knowledge", "User research" (46). Contrary to the practical investigations on users' behaviour (like questionnaires, protocols, or tests), lexicographers working

within the functional approach deduce the users' needs in collaboration with field experts. This task is accomplished by profiling the relevant characteristics of users, which are identified on the basis of pre-set general questions as well as on the direct experience of one professional who is an expert in the same specialised field of the dictionary. Additionally, the perspective situations (*Cognitive, Operative, Interpretative plus Communicative*) in which the dictionary is used allow the lexicographer to define the kind of data to be included in the reference work. A detailed exemplification of this "pre-compilation phase" for a specialised translation dictionary is offered in section 5.5. The necessary steps for transferring one text in one target language are carefully discussed, thus highlighting the many parts of this process in which monolingual instructions (both in the source and target language) are needed, such as a general introduction on the subject matter. In accordance with this analysis, the overall lexicographical design of dictionaries for specialised translations is outlined, using different component parts for L1-L2 translations and L2-L1 respectively. In addition, Fuentres-Olivera and Tarp discuss the dictionary part (e.g. the L1-L2 bilingual part or the monolingual section) in which different types of data should be included (e.g. collocations or grammar notes), since data distribution is another relevant topic regarding the reduction of information overload, which is a prerequisite for satisfying users' needs.

In fact, too much data make dictionary consultation inefficient and eventually make a search ineffective. This should be the major concern of lexicographers, as Fuentres-Olivera and Tarp argue, especially in the current Internet era, since the electronic environment offers solutions to tailor the entries according to the users' needs (97). For example, the tenets of function theory have guided researchers to compile "monofunctional" dictionaries, tools that are customised on the basis of the desired function, thus displaying different kinds of data if users need to write a text or, for example, understand a specific term. Displaying different kinds of entries is possible in the electronic environment, because data are stored in databases and retrieved by interfaces that filter them according to the specific consultation situation; e.g. inflected forms for writing and translating, "usage and/or contrastive notes" to lemma for a better understanding of terms (212). A detailed exemplification of the features of these innovative tools, corresponding to *Model T-Ford* in the evaluative classification of section 2.3, is given in the last chapter of the book.

The revolutionary aspect of these dictionaries is not limited to the employment of technological solutions, but to the achievements obtained using these new means, which allow users to select data more effectively, thus making it easier to transform them into information, a necessary cognitive process for a successful fulfilment of information voids. Thus, two interrelated factors are involved, "comprehension-related" and "search-related information costs" (142), which will also be used by the authors as assessment parameters in the critical overview of online specialised resources presented in Chapter 8.

Before dealing extensively with the actual solutions for editing a new class

of efficient dictionaries, in Chapter 9, Fuertes-Olivera and Tarp reflect more extensively on the lexicographical process, sketching their component parts both from the lexicographer's and the user's perspective.

While the first consists in three different compilation steps (pre-compilation, compilation, and post-compilation phase), the other is divided into an "extra-lexicographical pre-consultation phase", an "intra-lexicographical consultation phase", and an "extra-lexicographical post-consultation phase". Studies on dictionary use have mainly focused on the second, the intra-consultation phase; however the other two steps are considered to be more interesting by the authors, if they are approached focussing on how the user behaves "in the middle of these processes".

These researches should be carried out not by means of questionnaires and interviews, but using specific methods for observing users while they decide to "start a lexicographical consultation" (93), or testing the information they are able to retrieve from the dictionary. The first kind of investigation seems too costly to be carried out with statistically relevant demographic methods. The other, instead, is deemed more affordable and probably even more relevant, since in this phase it is possible to find the "proof of the pudding", namely evidence about the possibility of retrieving information from one dictionary, thus proving the quality of one dictionary as well as the "relevance of a lexicographical theory" (94).

One final remark should be made on another corollary of the theory of lexicographical functions, namely "proscription", or the way the lexicographer must behave with respect both to data documentation and user needs. In fact, while recording all possible forms and variants of one data, such as the many possible spellings of a particular word, the lexicographer should also signal (or recommend) the best choice available. This prevents the users from carrying out ineffective searches, which would occur if the nonstandard variant was not attested in the dictionary (this happens in prescriptive vocabularies), and allows them to receive the necessary advice about the correct form to use.

Chapter 6 (*Special Problems Related to Online Dictionaries*) starts with a complaint about the unimaginative use of the new powerful technologies, imputing the poor advances in lexicography to the "focus on practical problems related to natural language processing", as well as on other peripheral issues, such as user research. The pages of this chapter are instead devoted to showing how the new Internet environment can be exploited to obtain "higher lexicographical quality with lower production time and costs" (97).

The first type of technological devices discussed are the systems for data filtering, which allow compilers to offer tailored entries according to the consultation needs of users. The first option available is thus represented by the above-mentioned "monofunctional" access to specific kinds of data, which change on the basis of the type of consultation situation (e.g. writing, translating etc.). This option is adopted by many tools developed by the Centlex of Aarhus, as well as by the *Diccionarios de Contabilidad* presented in Chapter 9.

Filters, however, can tailor data even further, also taking into account the characteristics of the user, guiding him step by step with "interactive fill-in options" (99) to set the dictionary or, alternatively, giving him the chance to model the article by selecting the desired features.

The amount of data can be controlled also by adding information only when needed. Entries, for example, can initially display just a little information, such as indexes of contents or abstracts, as well as a few lines of text that can be expanded by means of "pop-up windows". The user may also be allowed to add his personal notes (e.g. synonyms) or, in other cases, additional material collected on the Web may be linked to one page, thus offering more information on one topic. In this respect, the authors are sceptical about the "automatic incorporation of data", which should always be validated by the lexicographer, in order to preserve the reliability of authored dictionaries. This involves a less optimistic view of possible future developments of lexicography, since the authors believe that only a semi-automatic generation of articles is possible. The entries may be pre-compiled by automatically retrieved materials, and then human experts should validate them. External and uncertified data should be explicitly signalled, and offered only to skilled users. It is obvious, however, that data from controlled corpora are more reliable than others that sprout up on the Internet with no restrictions.

6. Functions of *function theory*

The following sections of the book, Chapters 7 to 9, are real examples of how the theory can be productively used for different aims, such as evaluating other theoretical approaches or existing resources, as well as compiling real dictionaries. This part is thus a substantial integration of the first chapters, since the function theory is expected to be productive or, transformative, using the terminology adopted within this paradigm of studies. Therefore, the analysis of other competing models is one of the first concerns whilst presenting the actual application of function theory.

Chapter 7 offers, in fact, *A Critical View of Terminography*, reviewing five different theories belonging to this field of studies, from the General Theory of Terminology, promoted by Wüster, to the different knowledge engineering methods that deal with knowledge-based management systems. The aim of this revision is to prove how far these theoretical models are from the lexicographical approach supported by the authors, as well as the inadequacy of the terminological frameworks for a coherent lexicographical description of specialised languages.

The analysis is subsequently completed in the next chapter (*An Analysis of Specialised Online Dictionaries*), in which different specialised reference works are evaluated. Some of them have been compiled in accordance with the tenets of previously presented theories, such as the Communicative Theory of Terminology, by Cabré, which is the theoretical approach used for compiling the

Genoma project, or the frame-based model by Martin and Faber, which was used for the *EcoLexicon*. The revision process is carried out on the basis of ten assessment criteria, an "open list" that can also be improved with additional parameters. The declared goal of this section is to equip the confusing practice of "dictionary criticism" with principles that can turn it into a legitimate branch of lexicographical theory, as advocated by Wiegand and Gouws among others. This theoretical shift would allow both real users to receive valuable advice on the utility of existing dictionaries, and lexicographers to improve their works, referring to relevant analysis to their activities.

Results of this long enquiry (one of the longest sections of the whole book) are summarised in a table (201), assessing every evaluation feature for each dictionary. Additionally, the main findings are discussed in a brief list, which highlights that these specialised resources tend to accumulate irrelevant data for the envisaged functions, they are seldom updated, and are compiled without the assistance of any expert. A similar discredit to the characteristics of specialised languages is demonstrated by the pedestrian application of the same methodologies used for the analysis of the general languages. The major disappointment of the authors regards the use of corpora, which frequently serve as substitutes for the knowledge of experts, for example with respect to the lemma selection or the writing of definitions.

7. New tools in lexicography: examples from the *Accounting Dictionaries*

Contrary to these practices, the lexicographical process used for compiling the *Accounting Dictionaries* or *Diccionarios de Contabilidad* is noteworthy. The tool is a joint project of the lexicographers of the Aarhus University and of the University of Valladolid in Spain. Its title reveals that it is made up of a combination of different "monofunctional" dictionaries, assisting with cognitive and communicative tasks (211) all kinds of possible users: experts and semi-experts in the field of accounting, but also translators, students and laypersons (210). Chapter 9 (*Designing, Making and Updating Specialised Online Dictionaries*) is devoted to the description of the project in every phase, in order to give evidence of how theory becomes practice, and which methodologies make this process come true. The process is reported in detail, and offers many interesting topics for debate, also for specialised translation studies, which unfortunately cannot be dealt with here. Thus, the reader is recommended to go to these pages in order to have a real insight into a lexicographical process conceived within the tenets of function theory. Only a brief mention of some of the phases of the process and its main features can be made.

Firstly, the authors discuss the fact that the success of any lexicographical project relies on adequate planning of the software architecture, which is designed by IT experts and lexicographers together. Then users' profiles and the perspective situation of use of the dictionary allow the lexicographer to sketch the types of data to include.

Major problems of specialised dictionary projects may depend instead on the nature of the specialised language itself, like accounting terminology, which lacks conspicuous documentation and is not "very systematised" (214), since its jargon is used in daily working life and different accounting practices and rules are used all over the world and even within individual countries. Moreover, Spanish terminology is full of calques, loans and alternative possible translations of the English terms, while the language may also change with respect to different types of institutions, whether companies or non-profit organisations. The greatest concern of the *Accounting Dictionaries* was thus the selection of lemmas, which only in a small part were extracted from a corpus of three million words. The majority were instead selected from existing dictionaries, "institutionalised accounting texts", and "private texts" (215), using extensive reading, one assumes, since this aspect is not well explained. The reason for adopting this time-consuming procedure is that the habitual readers of these texts are also the prospective users of the dictionary, thus no other method seems adequate for the compilation of a lexicographical tool with sufficient terminology coverage. Additionally, the lemma list increases constantly and undergoes a continuous revision process, attesting new variants, synonyms or providing corrections. The measure of 500 new terms added in a period of eight months gives an idea of the ongoing nature of the work, something that is also possible because the dictionary is hosted in the digital environment and can be easily modified.

8. Final remarks

In conclusion, this book has only a few drawbacks, which do not call into question its relevance for lexicographical theory and practice. They regard the nature of the general lexicographical concepts that have been highlighted at the beginning of this review, with respect to the ideas of general language and general lexicography, in comparison with specialised language and specialised lexicography. These definitions have been proposed by the authors in order to make a brief abridgment to the field of specialised lexicography with a cursory definition, which would have required a more systematic analysis instead. The avoidance of referring to a broader reference science ('referensology' or 'accessology') has also made the dialectical continuity between the two sub-disciplines (general and specialised lexicography) probably less evident, whilst it has highlighted some residual part of the theoretical work that, in my opinion, remains to be done. The future research of lexicographical function theory, I believe, should deal also with the concepts of different types of languages (general and specialised) that lexicography describes, in order to become a complete, general theory of the discipline, not of a specific sub-field. Despite the fact that lexicographical theoretical principles remain the same, each language type (general and specialised) seems to entail different methodological solutions in order to satisfy the consultation needs of users. Therefore, the

overall lexicographical theory should probably deal more extensively with the specific requirements of these languages: their epistemic component (i.e. general and specialised knowledge), their variability (regional, national, local), and the type of documentation attesting them, as is demonstrated in the final part of the book on the *Accounting Dictionaries*.

Lastly, I have not touched upon many other topics, not because they are less important or less instructive, but because they would have derailed the discussion. I am still curious about some revision of old linguistic concepts by the lexicographical function theorists, such as 'collocation': "an umbrella term for referring to combinations of signs, typically words, which contextualise the meaning of lemmas and equivalents" (232). I would also have liked to read a more detailed discussion about definitions, not only describing what has been done in the *Accounting Dictionaries* (218-223), but also about the key topic of how definitions can facilitate users to turn the dictionary data into information, and what features may allow the lexicographer to accomplish this difficult task.

However, since the issues are correctly stated, many key lexicographical topics are therefore, at least, easier to address. This is also rendered possible by the use of new terminology, introduced by different theorists of the lexicographical function theory, which appears to fulfil its role well.

Notes

1. "Nor should it take its point of departure in the differences that separate all these works in terms of their specific content, structure etc., but in the aspects and elements that unite them and are common to all of them". (39)
2. The 'container' metaphor is my proposal, and it is not used in the book by Fuertes-Olivera and Tarp.

References

- Almind, R., H. Bergenholtz and V. Vrang.** 2006. Theoretical and Computational Solutions for Phraseological Lexicography. *Linguistik Online* 27(2). [http://www.linguistik-online.de/27_06/almind_et_al.html]
- Bejoint, H.** 2010. *The Lexicography of English*. Oxford: Oxford University Press.
- Bergenholtz, H. and R. Gouws.** 2010. A New Perspective on the Access Process Hermes. *Journal of Language and Communication Studies* 44: 103-127.
- Bergenholtz, H. and S. Tarp (Eds.).** 1995. *Manual of Specialised Lexicography: The Preparation of Specialised Dictionaries*. Amsterdam/Philadelphia: John Benjamins.
- Bothma, T.J.D. and S. Tarp.** 2012. Lexicography and the Relevance Criterion. *Lexikos* 22: 86-108.
- De Mauro, T.** 2004 (1980). *Guida all'uso delle parole: parlare e scrivere semplice e preciso per capire e farsi capire*. Rome: Editori riuniti.
- De Mauro, T. and I. Chiari.** 2014. Il vocabolario di base dell'italiano dal 1980 a oggi: metodi e risultati. The Sixteenth EURALEX International Congress ("The User in Focus"), EURALEX 2014,

- Bolzano/Bozen, Italy, July 15–19, 2014. Bolzano/Bozen: EURAC. Oral presentation.
- Fuertes-Olivera, P.A. and H. Bergenholtz.** 2011. Introduction: The Construction of Internet Dictionaries. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 1-16.
- Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.).** 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum.
- Fuertes-Olivera, P.A. and M. Niño-Amo.** 2011. Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés-Español de Contabilidad*. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 168-186.
- Gamper, J. and J. Knapp.** 2002. Towards an Adaptive Learners' Dictionary. Brusilovsky, P., O. Stock and C. Strappavara (Eds). 2002. *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2000), August 28–30, 2000, Trento, Italy*: 311-314. Berlin/New York: Springer.
- Henriksen, P.** 1992. Forholdet mellem encyklopedologi og leksikografi. Fjeld, R.V. (Ed.). 1992. *Nordiske Studier i Leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.–31. mai 1991*: 219-227. Oslo: Nordisk Forening for Leksikografi.
- Leroyer, P.** 2011. Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 121-140.
- Lew, R.** 2008. Lexicographic Functions and Pedagogical Lexicography: Some Critical Notes on Sven Tarp's Lexicography in the Borderland between Knowledge and Non-knowledge. Iwan, K. and I. Korpaczewska (Eds.). 2008. *Przegląd Humanistyczny. Pedagogika. Politologia. Filologia*: 114-123. Szczecin: Szczecińska Szkoła Wyższa Collegium Balticum.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer.
- Tarp, S.** 2011. Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. Fuertes-Olivera, P.A. and H. Bergenholtz (Eds.). 2011: 54-70.
- Tarp, S.** 2012. Online Dictionaries: Today and Tomorrow. *Lexicographica* 28(1): 253-268.

Evalueringskriteria en die interaksie tussen die leksikografieteorie en -praktyk; die ontwerp van die *Woordeboek van die Afrikaanse Taal* as gevallestudie

P.H. Swanepoel, *Departement Afrikaans en Algemene Literatuurwetenskap,
Unisa, Pretoria, Suid-Afrika (swaneph1@unisa.ac.za)*

Opsomming: Die samestellers van die verskillende dele van die WAT kon steun op die insette van verskeie medewerkers, waarvan woordeboekresensente 'n belangrike groep is. Een van die algemene probleme met die evaluering van woordeboeke van allerlei soorte (sien Nielsen 2006, 2009 en Swanepoel 2008, 2013 vir 'n bespreking) is dat daar egter in resensies — ook in sommige van dié van die verskeie dele van die WAT — slegs gesteun word op kriteria vir die ontleding en beskrywing van die ontwerpkenmerke van woordeboeke en dikwels nie ook op 'n duidelik geformuleerde en gemotiveerde stel kriteria vir die evaluering van woordeboeke nie. Hierdie probleem en die onderskeid tussen beskrywingskriteria en evalueringskriteria en die inhoud, struktuur en operasionalisering van evalueringskriteria word in meer besonderhede in hierdie artikel bespreek na aanleiding van veral Feinauer (2007), Gouws (2007, 2011 en 2014), Gouws en Prinsloo (2005), Nielsen (2006, 2009) en Svensén (2009). Ten slotte word hierdie probleem toegelig aan die hand van die ontwerpkenmerke van WAT Deel XIV.

Sleutelwoorde: WOORDEBOEKRESENSIES, WOORDEBOEKBEKRYWINGSKRITERIA, WOORDEBOEKEVALUERINGSKRITERIA, DIE VORM, INHOUD EN OPERASIONALISERING VAN LEKSIKOGRAFIESE EVALUERINGSKRITERIA, ONTWERPKENMERKE VAN WAT DEEL XIV

Abstract: **Evaluation criteria and the interaction between lexicographic theory and practice; the design of the *Woordeboek van die Afrikaanse Taal* as case study.** The compilers of the various volumes of the WAT have relied on input from various co-workers, of which dictionary reviewers form an important group. However, a common problem with the evaluation of dictionaries of various kinds (cf. Nielsen 2006, 2009 and Swanepoel 2008, 2013) is that reviews only provide an analysis and description of the design features of dictionaries and not also an evaluation based on a set of clearly formulated and motivated criteria for the evaluation of dictionaries. This problem, the distinction between criteria for the description and evaluation of the design features of dictionaries, and the content, structure and operationalization of lexicographic evaluation criteria, are discussed in this article, taking especially Feinauer (2007), Gouws (2007, 2011, 2014), Gouws and Prinsloo (2005), Nielsen (2006, 2009) and Svensén (2009) as

point of departure. Finally, this lexicographic problem is discussed with reference to the design features of WAT Volume XIV.

Keywords: DICTIONARY REVIEWS, DICTIONARY DESCRIPTION CRITERIA, DICTIONARY EVALUATION CRITERIA, THE FORM, CONTENT AND OPERATIONALIZATION OF LEXICOGRAPHIC EVALUATION CRITERIA, DESIGN FEATURES OF WAT, VOLUME XIV

Inleiding

Met die ontwerp van die verskillende dele van die *Woordeboek van die Afrikaanse Taal* (afgekort tot *WAT*) het die redaksie tot dusver kon profiteer van bydraes van verskeie ander medewerkers: taalliefhebbers, gebruikers van die *WAT*, kenners van verskillende vakgebiede, (Suid-)Afrikaanse taalkundiges, leksikografe, kenners van die leksikografieteorie en resensente wat in akademies geakkrediteerde media publiseer (sien Nielsen 2009: 208-210 en die bedankings in die voortekste van die verskeie dele van *WAT* en resensies soos dié van Boshoff 1926; Combrink 1962, 1979; Feinauer 1996, 2007; Gouws 1985, 1994; Grobler 1978; McLachlan 2010; en Odendal 1961a, b, 1962).

Resensente het egter die moeilike taak om nie net 'n beskrywing van die ontwerpprosesse en ontwerpkenmerke (soms verkort tot: *ontwerp*) van die verskillende (dele van) woordeboeke, soos die verskeie dele van die *WAT*, te gee nie, maar hulle moet ook 'n evaluering van die betrokke woordeboek(-dele) aanbied.

Een van die algemene probleme met die evaluering van woordeboeke van allerlei soorte (sien Swanepoel 2008 en 2013 vir 'n bespreking) is dat die voor-skrifte, wenke, matrikse of leidrade vir sulke resensies dikwels uit nie veel meer bestaan nie as 'n reeks rubrieke — gewoonlik 'n lysie van die makro- en die mikrostruktuurkenmerke — vir die ontleding van die ontwerpkenmerke van 'n woordeboek (van 'n bepaalde tipe). Kriteria vir die *evaluering* van hierdie kenmerke van woordeboeke word selde verskaf (sien Nielsen 2006, 2009 en Swanepoel 2013). Die gevolg is dat meeste resensies niks meer as 'n ontleding van die ontwerpkenmerke van 'n woordeboek is nie.

In ander gevalle word wel evaluering van die ontwerpkenmerke van 'n woordeboek verskaf, maar sulke uitsprake is dan gebaseer op die intuïsie of ervaring van die resensent in die rol van 'n enkele gebruiker van die betrokke woordeboek (of dan 'n enkele ingeligte gebruiker). Soms verwys resensente na die feit dat gebruikers van die woordeboek (heelwaarskynlik) baat sou vind deur alternatiewe of bestaande ontwerpkenmerke of -benaderings (byvoorbeeld ten opsigte van die vindbaarheid of verstaanbaarheid van data in 'n woordeboek), maar sonder dat hierdie soort evaluatiewe uitsprake deur 'n motivering onderbou word (byvoorbeeld motivering vanuit teoreties-gedrewe empiriese navorsing). Dit laat veel te wense oor as resensente eenvoudig hulle persoonlike oordeel oor die ontwerp van 'n woordeboek veralgemeen na alle potensiele teikengebruikers daarvan, of as enigste motivering vir evaluatiewe

uitsprake volstaan word met verwysings na die voorkeure, afkeure, woordeboek-gebruikersvaardighede, ensovoorts van teikengebruikers wat nie self empiries afgebaken is nie of as die gebruik van 'n woordeboek nie nagevors is nie.

Resensente het dikwels ook die neiging om skerp kritiek op bepaalde woordeboeke/-dele te lewer op basis van die oortreding van kriteria waarvoor daar geen duidelike motivering verskaf word nie, of waarvan dit nie presies aangedui word hoe resensente 'n bepaalde criterium vir die evaluering van die ontwerpkenmerke van 'n woordeboek moet operasionaliseer nie (sien byvoorbeeld Grobler 1978).

'n Verdere probleem met die beskrywing en evaluering van die WAT in die besonder is dat daar min beskikbare navorsingsliteratuur is oor die inhoud en struktuur en die gebruik al dan nie van beskrywingskriteria en evalueringkriteria in die resensies van verskillende dele van die WAT. Verder is dit nie duidelik of en hoe die samestellers van die WAT resensies soos dié wat bo vermeld word gebruik in die hersiening van verskeie dele van die WAT of die beplanning en samestelling van nuwe dele nie. Sien egter in hierdie verband Botha (2004, 2006).

Die eerste doelstelling van hierdie artikel is gevolglik om hierdie probleem van die evaluering van woordeboeke te bespreek aan die hand van die kriteria wat Svensén (2009: 480-487) vir 'n "goeie" leksikografiese resensie/woordeboekkritiek bied. Tweedens word hierdie probleem belig vanuit Gouws (2007, 2011 en 2014). In Gouws (2014) word 'n driedeling van die historiese ontwikkeling van die leksikografieteorie (en -praktyk) aangebied: die linguistiese benadering, die struktuurbenadering en die funksionele benadering, wat verskeie moontlikhede vir die ontwikkeling van beskrywings- en evalueringkriteria vir resensente bied. Dieselfde geld vir Gouws (2007) se voorstel vir die uitbreiding van die leksikografiese funksies vir 'n woordeboek na al die tekste daarvan (voortekste, agtertekste, mediotekste, sentrale woordelys, ensovoorts).

In hierdie artikel word daar dus doelbewus afgewyk van die navorsingsmetodologie oor resensies wat op basis van die inhoud van 'n korpus resensies (as't ware van onder na bo) empiriese data genereer vir die ontleding, beskrywing en evaluering van woordeboekresensies (sien byvoorbeeld Chan en Taylor 2001; Ripfel 1989).

Tweedens word daar 'n voorlopige raamwerk aangebied vir die ontleding en beoordeling van evalueringkriteria self (sien ook Swanepoel 2008 en 2013), alvorens 'n kort bespreking van die ontwerpkenmerke van WAT XIV verskaf word.

Volgens Botha (2004: 16) moet daar in die leksikografie 'n voordelige wisselwerking tussen die leksikografieteorie en -praktyk wees en dat die ontwerp van spesifieke woordeboeke "steeds die stand van die metaleksikografie (of leksikografieteorie — PHS) sal weerspieël". Daarmee het die WAT, by name van die huidige hoofredakteur, daarvoor 'n baie belangrike evalueringkriterium gestel, naamlik die mate waarin die redaksionele beleid en praktiese ontwerp van dele van die WAT aan die bestaande ontwikkelinge op die gebied van die

leksikografieteorie voldoen.

Die hoofdoel van hierdie artikel is nie om 'n volledige resensie van WAT Deel XIV aan te bied nie. Die fokus val eerstens op die problematiese aspekte van resensies van woordeboeke, en dié word ten slotte kortliks betrek op die ontwerpkenmerke van WAT Deel XIV.

Vereistes aan woordeboekresensies

Svensén (2009: 485) meld dat woordeboekresensies verskillende doelstellinge/funksies kan hê, waarvan die volgende vier die belangrikste is:

- Beskrywing
- Evaluering
- Motivering van/redes vir evaluering
- Raad/"Advice" (aan potensieële kopers en woordeboeksamestellers — PHS).

Svensén (2009: 482) wys dan daarop dat daar al bespiegel is oor watter van hierdie elemente as verpligtend vir 'n woordeboekresensie beskou moet word. Sy eie uitgangspunt is dat dit wel moontlik is om resensies te produseer wat slegs 'n beskrywing (van die ontwerpprosesse en ontwerpkenmerke) van 'n woordeboek bied, maar dat 'n goeie resensie (Engels: "regular review") ook 'n evaluering van hierdie kenmerke, 'n motivering vir elkeen van die gekose evalueringskriteria en advies vir potensieële gebruikers (en die samestellers van 'n woordeboek) moet verskaf.

Svensén (2009: 483) fokus veral op kriteria vir die beskrywing van die ontwerpkenmerke van 'n woordeboek, en hiervoor gooi hy die net breed uit deur nie minder nie as ses en dertig beskrywingsrubrieke vir die ontleding van die ontwerp van 'n eentalig verklarende woordeboek aan te bied. Die matriks word in vertaalde en aangepaste vorm hier onder aangebied omdat die rubrieke elkeen ook 'n aspek van die ontwerpkenmerke van 'n omvattende woordeboek soos die WAT dek.

1 woordeboekfunksies	19 grammatika
2 woordeboekgebruikers	20 ortografie
3 voorligting aan gebruikers	21 uitspraak
4 prys	22 semantiese en ensiklopediese inligting
5 uitleg/webontwerp	23 diasistemiese inligting
6 die samestellers	24 etimologie
7 vergelyking met ander woordeboeke	25 voorbeelde (sitate en poëme)
8 voorgeskiedenis van die woordeboek	26 kollokasies

9 verwysing na ander resensies	27 idiome
10 die resesent	28 illustrasies
11 woordeboekbasis/-dataversameling	29 sinonimie/antonimie
12 voortekste	30 kruisverwysings
13 lemmaseleksie	31 vermaaklikheidswaarde
14 bepaling van lemmas vir opname	32 opsommende/samevattende slotevaluering
15 soek- en toegangstrukture	33 tipes en graad van tekstuele verdigting
16 artikelstruktuur	34 makrostruktuur (seleksie en rangskikking van lemmas)
17 normatiewe/deskriptiewe dimensie	35 morfologiese inligting
18 ekwivalente	36 opname van seksuele, ras-, godsdienstige en nasionaliteitstereotipes

Alhoewel 'n resensie volgens Svensén (2009: 484) nie al hierdie aspekte van die ontwerp van 'n woordeboek moet dek om as "perfectly satisfactory" beskou te word nie, bied hy self nie duidelikheid oor hoe mens aan die hand van hierdie rubrieke 'n resensie van 'n woordeboek kan opstel en self evalueer na gelang van die graad van "bevrediging" ("satisfaction") wat 'n resensie bied nie (en, sou mens ook kon vra: Vir wie ook al?); byvoorbeeld, Svensén (2009: 452-479) gee self geen verdere leiding oor watter aspekte van die funksies en gebruikers van 'n bepaalde woordeboek, die voortekste, agtertekste of die makro- en mikrostrukturele kenmerke van die sentrale woordelys ontleed en bespreek moet word nie.

Rubrieke 33–36 neem Svensén (2009: 483) nie as deel van sy matriks op nie, maar noem dit wel as aspekte wat vir die ontleding en beskrywing van die ontwerp van 'n woordeboek geneem kan word. Sommige hiervan (byvoorbeeld 34 en 35) sou onder van die voorafgaande rubrieke bespreek kon word. Uiteraard sou meerdere ontwerpkenmerke bygevoeg kon word. 'n Duidelike leemte in hierdie matriks is byvoorbeeld die feit dat daar geen rubriek vir die ontleding en beskrywing van die agtertekste van 'n woordeboek is nie.

Gegee verder die gebrek aan aanvullende omskrywing van die rubrieke as beskrywingskriteria, is dit selfs moeilik om 'n matrys soos die bostaande te gebruik as kontrolemiddel vir die volledigheid van woordeboekresensies, of voortekste, soos die gebruikerstekste, of die sentrale woordelys van 'n woordeboek. Hiervoor sal 'n matriks soos die bostaande vir die ontleding en beskrywing van enige — veral die omvattende — woordeboek heelwat aanvulling moet kry. Resesente sou byvoorbeeld meer lig op die leksikografiese status van die rubrieke as beskrywingskriteria kon werp deur te gaan kyk wat leksikograwe in hulle redaksionele ontwerpbeleid oor elkeen sê, 'n studie maak van

die gebruikersinligting — indien volledig genoeg — as dit in 'n woordeboek (gewoonlik as voorteks) verskaf word, of die magdom navorsingsliteratuur oor die gebruik van elkeen van hierdie rubrieke as 'n kriterium vir die ontleding en beskrywing van ('n aspek van) 'n woordeboek. 'n Alternatiewe metode sou wees om die moeisame pad te volg om alle werke wat reeds as handboeke vir die samestelling van woordeboeke die lig gesien het, vir beskrywingskriteria (en evalueringskriteria) te myn (sien byvoorbeeld Atkins en Rundell 2008; Fontenelle 2008; Gouws en Prinsloo 2005; Hausmann 1989–1991; Svensén 2009; Van Sterkenburg 2003; Zgusta 1971).

Benewens die stel van die vereistes dat 'n resensie 'n evaluering van die ontwerpkenmerke van 'n woordeboek moet bied en 'n motivering vir die gekose evalueringskriteria moet verskaf, gee Svensén baie beperk aandag aan die ontwerp (inhoud, struktuur, vorm) en operasionalisering van elkeen van die leksikografiese evalueringskriteria en hulle motivering soos hulle op elkeen van die rubrieke in hierdie matriks betrekking het; byvoorbeeld: Daar word nie vermeld volgens watter kriteria die keuse van en inligting oor die gebruikers en gekose funksies van 'n woordeboek geëvalueer moet word nie, of watter motivering daar bestaan vir die keuse van spesifieke kriteria hiervoor nie. Svensén (2009) bied ook wat die evaluering van resensies betref, geen kruisverwysings na relevante dele van die teks van sy boek nie, maar volstaan met 'n klein handjie vol relevante navorsingsliteratuur oor woordeboekresensies (sien Svensén 2009: 487).

Svensén (2009: 485-486) keer egter na die onderwerp van die evaluering van woordeboeke terug onder die titel "Demands on a good dictionary review" en meld, veral na aanleiding van die beskouings van Nielsen (2003), enkele bykomende kriteria vir die evaluering van resensies self (Svensén 2009: 485-486). 'n Oorkoepelende vereiste van Nielsen is dat 'n resensie 'n regverdigde ("fair") weergawe van 'n woordeboek en die kenmerke daarvan moet gee. Om dit te bereik, word die volgende vereistes gestel aan 'n resensie:

- 'n Resensie moet net *relevante* inligting verskaf en evaluerings maak, dit wil sê, die inligting en evaluerings moet gekoppel wees aan die doelstellings van die woordeboek en bruikbaar wees vir die gekose teikengroep.
- 'n Resensie moet fokus op *essensiële* kenmerke van 'n woordeboek, dit wil sê dit moet slegs die kenmerke bespreek wat, indien hulle uit die resensie weggelaat sou word, 'n verkeerde indruk van die gehalte van die woordeboek kan skep.
- 'n Resensie moet *betroubaar* wees, dit wil sê dit moet nie verkeerde inligting bevat nie of steun op ongegronde stellings wat die potensiële gebruiker se opvatting van die woordeboek negatief kan beïnvloed.
- 'n Resensie moet *neutraal* wees deurdat dit alle relevante en essensiële uitgangspunte moet bevat, dit wil sê die resensent moet hom/haar nie deur haar/sy eie voorkeure in hierdie verband laat lei nie.

Svensén (2009: 485-486) voeg ook nog die volgende by:

- 'n Resensie moet nie in algemene terme geklee wees nie, maar fokus op spesifieke kenmerke van 'n woordeboek, soos die datakategorieë wat dit bevat.
- As 'n resensie na foute in die ontwerp van 'n woordeboek verwys (byvoorbeeld die weglating van lemmas), moet die resensent (byvoorbeeld by wyse van steekproewe) aandui hoe frekwent 'n afwyking/fout voorkom en hoe ernstig dit opgeneem moet word in die evaluering van 'n woordeboek (byvoorbeeld elke bladsy, sporadies deur die hele woordeboek, ensovoorts).
- In 'n resensie moet die slotevaluering die korrekte afleidings maak uit evaluering wat vroeër in die resensie gegee word, dit wil sê as 'n resensie die ontwerp van 'n woordeboek net/oorwegend as negatief beoordeel, dan kan die slotevaluering nie met 'n positiewe algemene evaluering eindig nie.

Uit Svensén (2009) is daar dus 'n stuk of elf kriteria vir die evaluering van die resensie van 'n woordeboek af te lei. Naas die feit dat 'n resensie 'n bespreking van die ontwerpkenmerke van 'n woordeboek moet bevat, moet dit ook 'n evaluering van hierdie kenmerke verskaf, 'n motivering bied vir elke evalueringskriterium en 'n slot-, opsommende evaluering van 'n woordeboek gee.

Relevante evalueringkriteria (hier geformuleer as vrae) — al is dit nie altyd duidelik oor presies hoe dié beantwoord moet word nie — is die volgende soos afgelei uit Nielsen (2003) hierbo:

- Bevat die resensie net *relevante* inligting?
- Fokus die resensie net op *essensiële* kenmerke van 'n woordeboek?
- Is die resensie *betroubaar*?
- Is die resensie *neutraal*?
- Fokus die resensie op spesifieke kenmerke van die ontwerp van 'n woordeboek?
- Word daar in 'n resensie ondersteuning uit die woordeboek self gebied vir die frekwensie en omvang van foute wat in die ontwerp van 'n woordeboek uitgewys word?
- Is die slotevaluering in 'n resensie korrek afgelei uit die positiewe of negatiewe evaluering wat vroeër in die resensie verskaf word?

Skakel mens dus Svensén (2009) se kriteria vir die skryf van 'n goeie resensie van 'n woordeboek om na 'n stel evalueringkriteria, dan is dit al moontlik om ten minste sommige aspekte van 'n resensie te beoordeel. Die kriteria is egter

vaag geformuleer en daar word nie presies aangedui hoe hulle vir die beoordeling van resensies geoperasionaliseer moet word nie. Hierdie punt word verder opgeneem hieronder.

Nielsen (2009: 211) meld verder dat resensente in hulle resensies 'n getroue ("faithful") weergawe van leksikografiese elemente van 'n woordeboek moet verskaf, en dat 'n getroue weergawe net moontlik is as alle teksdele van 'n betrokke woordeboek ontleed en beoordeel word; dus: ook aandag gee aan alle voor-, medio- en agtertekste en nie net die sentrale woordelys nie. Hierdie kriterium het in hoofsaak betrekking op die omvang van die woordeboek wat in 'n resensie aan die bod moet kom. As evalueringskriterium vir 'n resensie kan hierdie vereiste soos volg geformuleer word:

- Behandel en evalueer 'n resensie alle samestellende tekste van 'n woordeboek?

Nielsen (2006, 2009) wys byvoorbeeld daarop dat baie resensente eenvoudig in hulle resensies belangrike voortekste, mediotekste of agtertekste buite beskouing laat. Die beskrywing en evaluering van tekste wat (as voortekste) inligting verskaf oor die gebruik van woordeboeke (gebruiksgidse) word byvoorbeeld dikwels eenvoudig weggelaat of slegs op minimale wyse in resensies behandel.

Woordeboekontwerp en -evaluering en die interaksie met leksikografiese teorieë

Gouws (2014) tref 'n onderskeid tussen drie komplementerende fases in die ontwikkeling van die leksikografieteorie en -praktyk: die linguïstiese era waarin linguïstiese teorieë die basis vir die beplanning, ontwikkeling (en uiteindelijke evaluering) van 'n woordeboek was (sien veral Zgusta 1971); die era van woordeboekstruktuur waarin daar veral uitgebrei is op die struktuurkenmerke van woordeboeke en, maar in kleiner mate, die vereiste dat die ontwerp van 'n woordeboek deur die beoogde funksies en teikengebruikers daarvan bepaal moet word (sien veral Wiegand 1996; Gouws 2014); en die funksionalistiese benadering van die Arhauskool (Tarp en Bergenholtz; sien byvoorbeeld Tarp 2004a en b) waarvolgens die ontwerp en evaluering van 'n woordeboek bepaal moet word deur die linguïstiese en leksikografiese vaardigheide van die beoogde teikengebruikers daarvan en die funksie(s) waarvoor die woordeboek ingespan word in verskillende gebruikskontekste.

Gouws stel dit duidelik dat Zgusta (1971) die noodsaak bevestig van 'n linguïsties-teoretiese begroning van die ontwerp en uiteindelik evaluering van woordeboeke en vir 'n noue interaksie tussen leksikografieteorie en -praktyk (Gouws 2014: 379). Omgeset is daar hier sprake dat nie net die ontwerp en samestelling en die ontleding en beskrywing van die ontwerp van 'n woordeboek op basis van 'n duidelik en omvattende ('n) linguïstiese teorie(ë) moet geskied nie, maar ook vir die evaluering daarvan. In laasgenoemde geval kan

'n linguistiese teorie (as deel van omvattender leksikografieteorie) ook vir die evaluering en die motivering van die ontwerpkenmerke van 'n woordeboek dien; ter illustrasie: Die vertrekpunte van die tradisionele grammatika en die strukturalisme is in Zgusta (1971) 'n kernelement van die leksikografieteorie wat hy voorstaan. Hierdie teorie(ë) bepaal ook sy voorstelle vir die beplanning en ontwerp van 'n woordeboek (die seleksie, inhoud en strukturering van linguistiese data). Die evaluering van die ontwerpkenmerke van 'n woordeboek kan gevolglik gemeet word in dié mate waarin 'n woordeboek aan die vereistes van die betrokke teorie voldoen. Die kriteria wat hiervoor ingespan word, is gemotiveer vir sover daar empiriese ondersteuning vir die betrokke linguistiese teorie(ë) is. (Sien die bespreking hieronder.)

Wat Gouws (2014) oor die interaksie tussen woordeboekontwerp en die linguistiek/leksikografie hierbo maak, noop mens om weer Zgusta (1971) van vooraf en enduit te bestudeer. Hierdie werk laat sien hoe Zgusta linguistiese en leksikografieteorieë integreer om oplossings te bedink vir bepaalde leksikografiese probleme wat hulle voordoer in enige poging om 'n omvattende beskrywing, maar veral van die betekenisinhoud en -struktuur van die leksikale items (sien Zgusta 1971: 21-118) van 'n taal in 'n omvattende woordeboek te gee (sien Zgusta 1971: 222-293). In die geheel gesien, lewer die aangeduide dele van Zgusta 'n omvattende, linguisties gemotiveerde deskriptiewe raamwerk vir die ontwerp (sowel die ontwerpproses, as die ontwerpkenmerke), evaluering en die motivering van die kriteria vir laasgenoemde. (Sien ook Zgusta se bespreking van die leksikografiese hantering van meerwoordige leksikale items (1971: 138-163) en leksikale variëteit- en leksikale stylverandering (1971: 164-197) en die mikrostrukturele data-elemente wat ten opsigte van elke lemma verskaf moet word.)

Gouws (2011: 137) sluit indirek by laasgenoemde aan as hy dit stel dat die omvattendheid van 'n woordeboek bepaal word deur die omvang van die leksikale items wat as lemmas ingesluit word, die spektrum van die leksikon van 'n taal wat gedek word (byvoorbeeld die keuse van verskillende variëteite en dialekte), die verskeidenheid datatipes wat vir bewerking ingesluit word en die aard en omvang van hulle bewerking. Presies wat elkeen hiervan behoort te behels, kan slegs gespesifiseer word met 'n omvattende linguistiese teorie oor die leksikon van 'n taal (omvang, inhoud, struktuur, ens.) en op basis van een wat ook empiriese ondersteuning geniet.

Die linguisties teoretiese benadering is kenmerkend van die (sy dit ouer) eentalig verklarende woordeboeke — 'n feit wat volgens Nielsen (2006, 2009) veral daaruit spreek dat die gebruikstoeligting in meeste woordeboeke van hierdie soort fokus op die makrostruktuur (opname en strukturering van lemmas) en die mikrostruktuur daarvan, d.i. die linguistiese inhoud, struktuur, bewerking, ensovoorts in afsonderlike artikels van die sentrale woordelys. Sowel die ontwerp van hierdie soort woordeboek en die beskrywing en evaluering daarvan in resensies word dus op eksplisiete linguistiese teorieë gebaseer.

Die benadering van die Wiegand-era fokus veral op die ontwikkeling van

'n analitiese apparaat vir die ontleding van woordeboekstrukture soos wat hulle empiries in verskillende woordeboeksoorte gestalte kry. Bekende en nuwe terme wat hierop betrekking het, is byvoorbeeld *raamstruktuur*, *geïntegreerde* en *niegeïntegreerde voortekste* en *agtertekste*, *sentrale woordelys*, *dataverspreidingstruktuur*, *makro-* en *mikrostruktuur*, *artikelstruktuur*, *adresseringstruktuur*, *toegangstruktuur*, *mediostruktuur*, ensovoorts. (Sien Gouws 2014; Gouws en Prinsloo 2005; Wiegand 1996 vir 'n uitgebreide bespreking van hierdie terme.)

Die tweede stel evalueringskriteria blyk uit Gouws se bespreking van die Wiegand-era en die kriteria wat daar ter sake gebring word:

Die data (in 'n woordeboek — PHS) word aangebied sodat die teikengebruikers van die betrokke woordeboek die nodige inligting aan die aangebode data kan onttrek. 'n Optimale inligtingsonttrekking is slegs moontlik indien die teikengebruikers toegang tot die data kry en dit op 'n ondubbelsinnige manier kan interpreteer. Die verpakking speel ook 'n belangrike rol in die sukses van woordeboeke. ... Daar is ook toenemend aandag gegee aan die mate waarin die teikengebruiker en sy of haar behoeftes en naslaanvaardighede 'n invloed moet hê op die inhoud en aanbieding van elke woordeboek. (Gouws 2014: 379-380)

Hieruit blyk byvoorbeeld die volgende as kriteria wat vir die beoordeling van 'n woordeboek sou kon geld: *die mate waarin woordeboekgebruikers optimaal hulle inligtingsbehoefte uit die aangebode data kan onttrek, met as subevalueringskriteria: die mate waarin woordeboekgebruikers toegang tot die data (vir spesifieke funksies) verkry en die mate waarin woordeboekgebruikers die aangebode data op 'n ondubbelsinnige manier kan interpreteer. Daarnaas is daar nog kriteria soos die volgende: die mate waarin die verpakking van die data as suksesvol beskou kan word; die mate waarin die ontwerp van 'n woordeboek aan die naslaanvaardighede van die woordeboekgebruikers voldoen; die mate waarin 'n woordeboek spesifieke data aanbied vir elke funksie waarvoor die gebruikers 'n woordeboek kan gebruik.*

Kernprobleme met hierdie soort verhulde aanbieding van evalueringskriteria is dat die kriteria self nie eksplisiet as evalueringskriteria aangebied word nie, die kriteria self vaag geformuleer is (Hoe interpreteer en meet mens byvoorbeeld *die mate waarin...x?*) en dat daar geen duidelike riglyne is oor hoe 'n resesent in die beoordeling van 'n woordeboek elkeen van hierdie kriteria in die praktyk kan/moet toepas nie. Meer algemeen: Daar word nie aangedui hoe elkeen van die kriteria in die praktyk geoperasionaliseer moet word vir die evaluering van 'n woordeboek nie.

Die mate waarin 'n woordeboek aan enige van hierdie kriteria voldoen, is egter ook 'n empiriese vraag wat in die beoordeling van 'n woordeboek empiries getoets moet word. Waar sulke navorsing wel bestaan, lewer die resultate dikwels verrassings op. 'n Goeie voorbeeld hiervan is die empiriese toetsing van Cumming et al. (1994) van die aanname dat die volsin- leksikografiese definisie 'n beter soort definisie is as die sinstuk-leksikografiese definisie. Wat hulle navorsing toon, is egter dat daar geen beduidende verskil in produksietake is wat gebaseer is op die volsin- of sinstukdefinisies nie (sien verder Swanepoel 2000).

Die gebrek aan teoreties-gestuurde empiriese navorsing ter ondersteuning van leksikografiese evalueringskriteria blyk ook duidelik in die geval van die soort leksikografiese funksionalisme van Tarp en Bergenholtz (sien Gouws 2014). Hiervolgens moet die ontwerp van 'n woordeboek (inhoud, struktuur, ens.) bepaal word deur die funksies (kommunikatief, kognitief, operasioneel en interpretatief) waarvoor spesifieke kategorieë taalgebruikers (met inagneming van hulle taalvaardigheid en naslaanvaardigheid) die woordeboek in spesifieke gebruikskontekste vir bepaalde inligtingsbehoefte inspan.

Die kommunikatiewe funksies word byvoorbeeld verder verdeel in enkoderings- en dekodeerstake. Die outeurs bied op basis van hulle eie kennis of intuïsie van wat elkeen van hierdie take aan data op die makro- en mikrostrukturele vlak van 'n woordeboek vereis (sien Tarp 2004a, b). Geen empiriese navorsing word ingespan om inderdaad te bepaal watter leksikografiese data spesifieke woordeboekgebruikers vir watter funksies in bepaalde gebruikskontekste inspan nie. Trouens, een van die kenmerke van hierdie vorm van leksikografiese funksionalisme is dat die meeste empiriese leksikografiese navorsing as oorbodig vir óf die ontwerp óf die evaluering van woordeboeke verklaar word. (Sien byvoorbeeld Tarp 2009.)

Nielsen (2006) volg 'n verwante, maar tog verskillende benadering. Aanvanklik beklemtoon hy die noodsaak vir empiriese navorsing oor woordeboekfunksies, -gebruikers en gebruikskontekste (Nielsen 2006: 3) vir die ontwerp van woordeboeke, maar by gebrek aan hierdie soort navorsing slaan hy al hoe meer oor na die modale kwalifisering van sy bespreking van die ontwerp en evaluering van woordeboeke met die woordjie "may". (Sien byvoorbeeld Nielsen 2006: 5 se uitlating dat "dictionaries may be multifunctional".)

Hierdie benadering tot die ontwerp en evaluering van woordeboeke steun metodologies op die deduktiewisme (en nie die induktiewisme nie), dit wil sê die ontwerp en evaluering van 'n woordeboek word nie ondersteun deur teoreties gedrewe empiriese leksikografiese navorsings nie, maar daar word vanuit die intuïsie, kennis of ervaring van die leksikograaf 'n teorie geponeer vir sowel die ontwerp as evaluering van 'n woordeboek. Hierdie soort funksionalisme is gevolglik ook niks meer as 'n teorie nie wat om empiriese toetsing vra alvorens die status van die ontwerp- en evalueringskriteria wat mens daaruit sou kon aflei, beoordeel kan word.

Gegee dat soveel idees oor die ontwerp en evaluering van woordeboeke op basis van funksies gebaseer is op ingeligte opinie, meld Swanepoel (2000: 405) die volgende:

Given that current thinking on the FQ (=Functional quality; PHS) of dictionaries is for a great deal based on personal experience, I will first tackle the status of subjective experience as a source of our knowledge on the complexities of the FQ of dictionaries.

Although lexicographical and pedagogical experience or "informed opinion" can guide hypothesis formation in theory construction and empirical testing, alone it is not enough to elucidate the real problems learners experience with the functionality of dictionaries. For one, these assumptions may simply be wrong,

misguided or have little empirical support beyond the limits of the lexicographer's, teacher's or even dictionary reviewer's own subjective experience ...).

Wat egter ook duidelik is, is dat die deduktivistiese benadering van Tarp en Bergenholtz oor die funksionele ontwerp van woordeboeke 'n baie belangrike bydrae gelewer het en nog steeds lewer tot die ontwikkeling van die leksikografieteorie, leksikografiese praktyk en die evalueringskriteria van woordeboeke — al vra hierdie bydraes om verdere empiriese toetsing.

'n Belangrike bydrae tot die ontwikkeling van woordeboekfunksies is egter Gouws (2007) (sien ook Nielsen 2006, 2009), wat die funksionele ontwerp en evaluering van woordeboeke uitbrei van die sentrale woordlys na die buitekste (voor- en agtertekste) toe. Die eerste evalueringskriterium wat hy stel, is dat 'n woordeboek in die gebruikerstoeligtig (as voorteks) die gebruiker moet inlig van al die strukture en funksies van die woordeboek as geheel en hoe die gekose strukture en funksies die ontwerp van al die samestellende tekste van 'n bepaalde woordeboek bepaal. Die noodsaak hiervoor, en veral dat daar vir woordeboekgebruikers 'n omvattende, maklik toeganklike gebruiksgids as voorteks gegee moet word en dat hulle deeglik daarvan moet kennis neem, word daarin gemotiveer dat sonder hierdie kennis gebruikers moeilik die data in die res van 'n woordeboek sal kan ontsluit (sien Gouws 2007: 80-81). Gebruiksgidse wat wel so ontwerp is, is geïntegreerde tekste vir sover hulle gebruikers help om die relevante inligting te vind vir die uitvoering van die funksie(s) wat die woordeboek in sy ontwerp in die oog het.

Gouws (2007: 84-85) gee as voorbeelde van agtertekste wat wel 'n duidelike funksie het maar waarvan die funksie van die sentrale teks kan verskil, 'n lys van vleissnitte met illustrasies en gedetailleerde (geografiese) kaarte. Hierdie agtertekste het duidelik primêr 'n kognitiewe funksie (uitbreiding van die kennis van die gebruiker) en dié kan funksioneel verskil van 'n sentrale lys wat primêr 'n kommunikatiewe funksie (enkodering en dekodeer van tekste) het.

Nielsen (2006, 2009) fokus ook op die funksionele ontwerp van woordeboeke en veral die gebruikstekste, en stel daarmee ook 'n aantal kriteria vir die evaluering van funksionele gebruikstekste. Volgens Nielsen (2009: 215) het 'n funksioneel ontwerpte woordeboek die volgende kenmerke:

- Dit is ontwerp om een of meer funksies te ondersteun.
- Die opgenoemde data is geselekteer omdat hulle spesifieke woordeboekfunksies ondersteun.
- Struktureel is 'n woordeboek so ontwerp dat die data die gekose funksies kan ondersteun.

In die bespreking van die ontwerp van gebruikstekste onderskei Nielsen (2006) duidelik tussen woordeboeke waarvan die ontwerp deur een of meer linguïstiese teorieë bepaal word en dié wat deur die funksionele benadering bepaal is. Eersgenoemde fokus in die gebruikstekste op die linguïstiese inhoud en struk-

tuur van die betrokke artikels van die sentrale woordelys. Funksioneel ontwerpte gebruiksgids, daarenteen, gee 'n uiteensetting van die verskillende maniere waarop die gebruiker die woordeboek kan gebruik om bepaalde funksies (spesifieke inligtingsbehoefte in spesifieke gebruikskontekste) uit te voer. Ontwerpsgewyse stel Nielsen (2006) voor dat 'n funksioneel ontwerpte gebruiksgids uit twee of meer dele kan bestaan:

- 'n visuele gids (skematiese voorstelling) van die data-inhoud en -struktuur in artikels
- 'n teks wat uit onderskeibare gedeeltes bestaan waarin in elke deel duidelike riglyne (instruksies en/of direktiewe) verskaf word vir die uitvoering van 'n bepaalde funksie met die hulp van die woordeboek
- 'n teks wat op basis van die linguïstiese benadering 'n uiteensetting gee van die makrostruktuur en van die linguïstiese inhoud en struktuur van die artikels

Soos hierbo uiteengesit, is Nielsen (2006, 2009) se voorstel dus dat die linguïstiese en die funksionele benaderings mekaar kan komplementeer.

Nielsen (2009: 218) verskaf 'n duidelike aantal funksioneel georiënteerde riglyne vir resensente vir die evaluering van die voortekste, en in die besonder gebruikstekste, van woordeboeke:

- die mate waarin die beoogde funksies (teksresepsie, teksproduksie, ensovoorts) uiteengesit word
- die mate waarin rekenskap gegee word van die omvang van die woordeboek (byvoorbeeld linguïsties, vakspesifiek, ensovoorts)
- die mate waarin die teikengebruikers van die woordeboek aangedui en afgebaken word
- die mate waarin die empiriese basis van die woordeboek uiteengesit word
- (meer spesifiek gerig op die evaluering van die gebruikstekste) of die gebruiksgids aandui hoe gebruikers die woordeboek moet gebruik vir elkeen van die beoogde funksies daarvan
- of die gebruikersgids aandui watter data in die woordeboek voorkom
- of die gebruikersgids aandui waar in die woordeboek elke datatipe gevind kan word
- of aangedui word hoe data in die buitetekste verband hou met of komplementêrend optree ten opsigte van die data in die sentrale woordelys en ander buitetekste

- of aangedui word in watter gebruiksituasie die opgenome data *waarskynlik* (indien nie empiriese navorsing hieroor gedoen is nie — PHS) sal kan help

'n Raamwerk vir die ontleding, beskrywing, beoordeling en motivering van leksikografiese evalueringskriteria

Die bespreking van die leksikografiese probleem hierbo noodsaak dat daar vir die skryf van 'n resensie 'n duidelike onderskeid getref moet word tussen ten minste 'n analities of deskriptiewe term (byvoorbeeld *raamstruktuur*, *makrostruktuur*, *lemma* en *uitspraak*) en 'n evalueringskriterium, dit is, breed gedefinieer, enige standaard, norm, beginsel, praktiese norm of reël of prosedure op basis waarvan die ontwerp-elemente van 'n woordeboek (as negatief of positief) beoordeel word.

Wat laasgenoemde betref, is dit verder nodig om te onderskei tussen

- die aspek van die ontwerp van 'n woordeboek waarop 'n evalueringskriterium betrekking het (byvoorbeeld die struktuur van 'n woordeboekartikel of leksikografiese definisies/betekenisverklarings)
- die inhoud en struktuur van 'n evalueringskriterium (byvoorbeeld: definisies/betekenisverklarings moet vir die gebruiker verstaanbaar wees. (stelling); Is die definisies vir die gebruiker verstaanbaar? (vraag)), of dat die gebruiker uit die aangebode data in 'n woordeboek die nodige inligting moet kan ontsluit wat hy of sy vir 'n bepaalde funksie in 'n bepaalde konteks nodig het)
- hoe evalueringskriteria geoperasionaliseer of toegepas kan word in die evaluering van 'n ontwerpkenmerk van 'n woordeboek (byvoorbeeld, hoe kan bepaal word of definisies deur die gebruikers verstaan word, of dat hulle inderdaad uit die aangebode data in 'n woordeboek die nodige inligting vir 'n bepaalde funksie suksesvol ontsluit het/kon ontsluit)
- watter motivering daar vir 'n bepaalde evalueringskriterium aangebied (kan) word (byvoorbeeld motivering vanuit 'n aspek van die leksikografieteorie, teoreties-gestuurde empiriese navorsing oor woordeboekgebruik, standaarde vir die ontwerp van bepaalde woordeboeke, gevestigde leksikografiese gebruike, ensovoorts.)

Die volgende aanvullende (maar deels oorvleuelende) raamwerk word verskaf vir die ontleding en beoordeling van die evalueringskriteria wat resensente van woordeboeke en leksikograwe kan inspan — hier met aanpassings en toeligende voorbeelde) (sien Swanepoel 2013: 593). Die stippellyne dui aan dat die evalueringskategorieë en evalueringskenmerke self nog oop is en om invulling met verdere navorsing vra.

Inligting wat deur die evalueringskriterium gedek word:

(a) Strekking:

Algemeen (geldig vir beoordeling van alle soorte woordeboeke)

Genre-spesifiek (byvoorbeeld, geldig vir eentalige aanleerderswoordeboeke)

Tekssoort-spesifiek (byvoorbeeld, het betrekking op verskillende soorte voortekste, die sentrale woordelys of die agtertekste)

Ontwerpkenmerk-spesifiek (byvoorbeeld, gemik op insluiting, inhoud en struktuur van morfologiese inligting)

(b) Omvattendheid:

Uitgebreide — Gebrekkige detail; arbitrêr (byvoorbeeld, dek die evaluering van alle morfologiese inligting in 'n artikel of slegs (en vaagweg) gemik op die evaluering van inligting oor afleidings)

Geldigheid/gemotiveerdheid van die evalueringkriterium:

Basis van motivering:

Standaard

Teorie

Gebruikersnavorsing (insluitende oor taalvaardigheid, leksikografiese naslaanvaardighede, gebruikskontekste en inligtingsbehoefte van gebruikers)

Leksikografiese praktyk/tradisie

Funksies van die woordeboek

Ontwerpkenmerke van spesifieke woordeboektypes

Vindbaarheid en verstaanbaarheid van inligting (van 'n bepaalde tipe) (byvoorbeeld leksikale definisies)

.....

Nuutheid/uitsonderlikheid van die kriterium:

Hoog — Laag

Aanbiedingsformaat/-struktuur van die evalueringskriterium:

Instruksie

Vraag

Stelling

Vereistes

Verduideliking
Toepassing van die evalueringskriterium: (a) Vlak van kundigheid wat vir die toepassing van die evalueringskriterium vereis word: Meganisties toepasbaar (byvoorbeeld, meting van aantal lemmas in 'n woordeboek) Vakkundigheid vereis vir toepassing (kan slegs toegepas word deur 'n leksikograaf, taalkundige, resensent, ensovoorts; byvoorbeeld, bepaling of definisies verstaanbaar is) (b) Uitkomste: Oop — geslote (enige uitkomste of beperkte of vasgestelde uitkomste van die toepassing van 'n evalueringskriterium) (c) Operasionalisering: (In detail) gespesifiseer — ongespesifiseer
.....

Alhoewel nie volledig nie, is hierdie ontleding 'n voorlopige poging om van die volgende vrae oor die gebruik van leksikografiese evalueringskriteria vir die beoordeling van 'n omvattende woordeboek, soos die dele van die WAT, te beantwoord:

- Watter kriteria word in die beoordeling van die ontwerpkenmerke van 'n omvattende, sinchronies verklarende woordeboek gebruik?
- Wat is die status van die evalueringskriteria, of hoe geldig is hulle vir die beoordeling van 'n omvattende verklarende woordeboek? Bestaan daar 'n goeie motivering vir die evalueringskriteria?
- Kan die evalueringskriteria geoperasionaliseer word of is hulle toepassing afhanklik van die intuïsie van 'n resensent?
- Kan die evalueringskriteria omskep word in praktiese riglyne vir die ontwerp van omvattende verklarende woordeboeke?
- Is daar empiriese bewyse daarvoor dat die toepassing van die evalueringskriteria (geformuleer as ontwerpheuristieke) die funksionaliteit van nuwe woordeboeke of die hersiening van bestaande woordeboeke sal verhoog?

Die bostaande is spesifiek op die beskrywing en beoordeling van evalueringskriteria gemik. Intussen het Feinauer (2007) met die beskrywing en evaluering van WAT Deel XII in haar resensie ook 'n sistematiese raamwerk vir die orde-

ning van hierdie kriteria en hulle toepassing gestel deur die raamstruktuur van 'n woordeboek hiervoor te benut, naamlik beskrywing en beoordeling van die voortekste, die sentrale woordelys en die agtertekste van 'n woordeboek. (Sien ook Gouws en Prinsloo (2005) se bespreking van die raamstruktuur van woordeboeke.)

Beskrywing en evaluering van die ontwerpkenmerke van WAT Deel XIV

WAT Deel XIV is gepubliseer in 2013 met dr. W.F. Botha as hoofredakteur. Naas hierdie deel, word nog twee dele beplan vir die voltooiing van die letter S.

Die ontwerp van die raamstruktuur van WAT Deel XIV volg dié van WAT Deel XIII: vier voortekste (*Inleiding, Skematiese voorstelling van inskrywings, Toeligting by die gebruik van die Woordeboek en Redaksionele afkortinge*, pp. I-XXI), die sentrale woordelys (wat die Afrikaanse woordeskat dek van *S-Skooi*, pp. 1-637) en twee agtertekste (*Bronne waaruit in hierdie deel van die Woordeboek aangehaal is*, pp. 638-672, en *Verbeterings in Deel XII van die Woordeboek*, pp. 673-674).

Hieronder word eers 'n aantal aspekte van die buitetekste bespreek alvorens die sentrale woordelys aan bod kom. As beskrywings- en evalueringsraamwerk word hier op die voorafgaande bespreking gesteun sonder dat alle fasette van hierdie raamwerk weer in besonderhede bespreek word.

Voortekste

Die afdeling *Inleiding* bevat hoofsaaklik inligting wat nie met die sentrale woordelys geïntegreer is nie. Sake soos die befondsing van die WAT, publikasies wat sedert 2009 verskyn het, digitalisering van die WAT se materiaalversameling, personeelsake, vakkundige medewerkers, 'n versoek om inligting en 'n versoek vir materiaalbydraes en kommentaar kom aan bod.

Wat opvallend in hierdie teks vanuit 'n evalueringsoogpunt is, is wel die (kort) missiestelling van die WAT (p. V) — 'n tipiese illustrasie van hoe vaag missiestellings kan wees:

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.

Uit hierdie stelling kan twee kriteria vir die evaluering van hierdie deel van die WAT afgelei word, naamlik spesifisering van die teikengroep en die beoogde funksies van hierdie deel van die WAT. Hieruit blyk in die besonder dat die WAT 'n veel breër groep gebruikers (of miskien meer vaag geformuleer: belanghebbendes) in die oog het as wat vermeld word in die eerste paragraaf van die *Toeligting by die gebruik van die woordeboek*, p. VIII ("alle sprekers van Afrikaans ... maar veral die spektrum gebruikers wat strek vanaf senior hoërskoolleerders tot Afrikaansakademië"). Volgens die missie is die teikengroep

alle gebruikers van Afrikaans, Afrikaans as taal, die (beoefenaars van die) leksikografie (-teorie en -praktyk?) op internasionale vlak en Suid-Afrikaanse beoefenaars van die leksikografie (-teorie en -praktyk?). Soos dikwels gebruiklik, is die doelstelling/missie gepas vaag gehou vir sover *dien* 'n multisemantiese leksikale item is waarvan die presiese betekenis (doelstellings/funksie) nie omskryf word nie. Dit is dus moeilik om uit die missiestelling 'n gemotiveerde evalueringskriterium vir die WAT (of 'n deel daarvan) af te lei. Hoe bepaal mens byvoorbeeld of ('n deel van die) WAT die gebruikers van Afrikaans of die internasionale of Suid-Afrikaanse leksikografie dien en wat *dien* in hierdie konteks sou beteken?

Feinauer (2007: 259) interpreteer die missie van die WAT egter in haar resensie as sou die missiestelling betrekking hê op die evaluering van bepaalde ontwerpkenmerke van WAT, Deel XII. In die opsomming van haar artikel maak sy die uitspraak dat die buitetekste met die sentrale teks geïntegreer is "en in daardie opsig dus wel die leksikografie dien". Verder is die slotsom dat die gebruikers van die Afrikaanse taal sowel as die Suid-Afrikaanse leksikografie "meer daarby sal baat" as die sentrale teks "deegliker gekonseptualiseer en beplan sou word aan die hand van leksikografiese funksies". As verdere tekens van hoe die WAT dan wel sy "dienstaak" suksesvol sou kon uitvoer, word die volgende genoem: opname van leksikale items moet aan die hand van 'n goed beplande korpus herbedink word, frekvent gebruikte woorde moet nie by lemmaeuses oor die hoof gesien word nie, die lengte van artikels moet gelykvormig gebalanseer word en 'n gebalanseerde verteenwoordiging van makro- en mikrostruktuur, wat ook daartoe sal lei dat daar minder onbewerkte lemma-tekens is.

Daar word hier volstaan met die opmerking dat dit nie duidelik is hoe *dien* en die relevante evalueringskriteria in mekaar steek nie. Weliswaar kan daar verskillende evalueringskriteria vir die WAT uit Feinauer (2007) se opsomming afgelei word, al word nie een van hulle onafhanklik gemotiveer nie. Miskien is die eerste probleem dat die WAT met so 'n vae missiestelling volstaan en nie presies duidelikheid gee oor wat *dien* in hierdie konteks beteken nie. Duidelik het resensente en leksikografe hier leiding nodig oor die funksie, inhoud en struktuur van 'n missiestelling en 'n duidelike beskrywing van die evaluering daarvan.

'n Tweede probleem is die feit dat daar nie in die afdeling *Toeligting by die gebruik van die Woordeboek* 'n verdere presiese afbakening van die teikengebruikers en die beoogde funksies van die WAT gegee word nie, of presies hoe teikengebruikers en funksies die ontwerpkenmerke van 'n deel van die WAT bepaal nie. Alle gebruikers van Afrikaans, hoërskoolleerders, en Afrikaansakademië is 'n besonder heterogene groep wat groot verskille sou kon toon ten opsigte van hulle eie kennis van Afrikaans, woordeboeknaslaanvaardighede en die funksies waarvoor hulle die WAT in bepaalde gebruikskontekste sou wou raadpleeg. Dit is nie te voorspel watter radikale veranderinge dit aan die ontwerp van die dele van die WAT kon impliseer nie. 'n Mens vermoed egter dat indien die samestellers die taak van die WAT sien as om die ganse woordeskat

van Afrikaans in al sy verskyningsvorme te *dokumenteer* (sien Botha 2006), dan is 'n deskriptiewe funksie heelwaarskynlik die primêre een. So 'n uitgangspunt sou ook strook met die struktureel deskriptiewe benadering wat tot die inhoud en struktuur van die artikels in die sentrale woordelys ingeneem word. Dié kan mens aflees uit die skematiese voorstelling van inskrywings (pp. VI-VII), al die datakategorieë wat in die toeligting by die gebruik van die woordeboek (pp. VIII-XXI) bespreek word en die datakategorieë wat in elke artikel aan die bod kom. Dit sou ook strook met 'n bron soos Zgusta (1971) wat uitvoerig ingaan op die taalkundige begroning van die omvattende (sinchroniese) woordeboek. As sodanige is WAT Deel XIV 'n baie deeglike omvattende sinchroniese woordeboek, waarvan net die kenmerke self in die toeligting by die gebruik van die woordeboek sekuur omskryf word (sien p. VIII).

Dit hou egter nie in dat die woordeboek nie wel 'n multifunksionele gebruik beoog nie. Tarp en andere het reeds lankal daarop gewys dat mens nie in die geval van gedrukte woordeboeke kan verwag dat daar vir elke moontlike funksie 'n monofunksioneel ontwerpte woordeboek beplan en uitgegee moet word nie (sien Bergenholtz en Bothma 2011: 60). Gegee die talle funksies waarvoor woordeboeke gebruik sou kon word, sal dit tot 'n (onbekostigbare) proliferasie van woordeboeke kan lei. Aan die ander kant word daarmee nie ontken dat die samestellers wel deeglik in die toeligting tot die gebruik van woordeboeke wel vir gebruikers touwys moet maak oor hoe 'n woordeboek soos die (verskillende dele van die) WAT vir bepaalde, goed omskrewe funksies benut kan word nie. Die breër probleem wat hierdie saak aan die orde stel is: Wat verwag gebruikers en leksikograwe van 'n "goeie" gebruiksgids? Aan watter vereistes moet dit voldoen en hoe evalueer mens dit? Dit is 'n onderwerp wat selde aan die orde gestel word. (Sien in hierdie verband egter Gouws 2007; Nielsen 2006, 2009.)

Aanvaar 'n mens dat die praktiese leksikografie deur ontwikkelinge op die gebied van die leksikografieteorie gevoed moet word (sien Botha 2004), dan sou 'n mens wel kon verwag dat die samestellers van die WAT van die sake wat hierbo genoem word, rekenskap moet gee, sy dit in die WAT self of enige ander gepubliseerde bron.

Vir eie gebruik, het ek die toeligting by die gebruik van die woordeboek te lank en ingewikkeld gevind. Alhoewel volledig, sou mens vir groter gebruiksvriendelikheid kon vra. Indien uit navorsing gemotiveer, sou mens graag wou sien dat die herskrywing van die inleiding, in die besonder inligting oor die gebruik van die woordeboek (of dan dele daarvan), in die toekoms aandag kry.

Agtertekste

Die teks onder *Bronne waaruit in hierdie deel van die Woordeboek aangehaal is* (pp. 638-672), moet in twee verdeel word: 'n afdeling *Toeligting* (p. 638), waarin die redaksionele beleid ten opsigte van die opname van bronne en verwysings na bronne op 'n baie insigtelike wyse uiteengesit word, al word daar in hierdie teks geen motivering vir die gekose werkswyse gebied nie. Die tweede deel

(pp. 639-672) bestaan uit 'n omvangryke lys van die gekose bronne in ooreenstemming met die beleid wat in die inleidende paragrafe onder die hofie *Toe-ligting* uiteengesit word. Laasgenoemde is dus geïntegreer met die res van die bronnelys en dié weer met die verwysings (sitate) wat as voorbeeldmateriaal in die artikels van die sentrale woordelys voorkom. In die geheel gesien, dra hierdie teks dus by tot die integrasie van die tekste van hierdie deel van die WAT. Hierdie evalueringskriterium het dus betrekking op die inhoud, struktuur en funksie van die verskaffing van die bronne waaruit daar in die sentrale woordelys vir verwysings (sitate in die besonder) gesteun word. Die motivering van hierdie gebruikswyse van bronne ontbreek, en aangesien alternatiewe verwysingstrukture en bronverwysings moontlik is, is dit opvallend dat daar vir die verkose werkwyse geen motivering gegee word nie.

Die teks *Verbeterings in Deel XIII van die Woordeboek*, in WAT Deel XIV, pp. 673-674, is 'n baie kort teks van twee bladsye met regstellings en verbeterings op die sentrale deel van WAT Deel XIII. Feinauer (2007: 264) het alreeds met 'n resensie van WAT Deel XII daarop gewys dat so 'n teks met verbetering en regstellings in WAT Deel XI nie met die ander tekste van WAT Deel XII te make het nie. Dieselfde geld die insluiting van 'n lys met regstellings aan WAT Deel XIII in Deel XIV. Die probleem met hierdie werkwyse is natuurlik dat die gebruikers nooit in 'n vorige deel alreeds en duidelik vertrouwd gemaak word met hierdie aanbieding van regstellings nie sodat hulle sal weet dat wat daar ook al fout met 'n deel is (weglatings, drukfoute, inkonsekwentheid in redaksionele beleid, opname van onverklaarde lemmas, ensovoorts) in 'n daaropvolgende deel gepubliseer sal word. Uiteraard kan die regstellings deels gemotiveer word op basis van resensies, leksikografiese literatuur of suiwer op basis van die proefleeswerk van die redakteurs van die WAT. Miskien sal gebruikers baat vind by 'n inleiding wat help om die artikels van hierdie veranderinge te verduidelik. Of gebruikers so 'n inleiding inderdaad nodig ag en nut daarvan sal hê, is 'n empiriese vraag wat empiriese navorsing vereis.

Dieselfde geld ten opsigte van twee ander oorwegings wat al twee die implikasie het dat die omvang van elke nuwe deel van die WAT sal vergroot, en, bykomend, of dit as geregverdig/gemotiveerd beskou kan word as mens dit opweeg teen die addisionele koste, personeelbeperkinge maar veral die poging om die WAT te voltooi.

Feinauer (2007: 264) lewer byvoorbeeld kommentaar op die feit dat die agtertekste nie benut word vir geïntegreerde tekste van inligtingskomponente wat verspreid in 'n woordeboekdeel (in aparte artikels of as deel van verskillende artikels) voorkom nie (byvoorbeeld 'n geïntegreerde lys met afkortings of met etimologiese inligting). Die vraag is natuurlik waar die woordeboekmakers op gemotiveerde wyse die grens sou moes trek met die opname in aparte tekste van hierdie soort inligting — daar kom immers talle datatipes in elke artikel voor. Samestellers van hierdie soort woordeboek het natuurlik ook die taak om te keer dat die omvang van enige deel van die WAT buite koste, hanteerbaarheid van inligting, ensovoorts eskaleer. Trouens, met die hersienings van die redaksionele beleid van die WAT is daar juis besluit om af te skaal wat die

inhoud van die WAT betref, soos byvoorbeeld die opname van vakterme en illustrasies.

Feinauer (2007: 264) wys verder daarop — en dit sou ook geld ten opsigte van WAT XIV — dat agtertekste nie benut word om nog bykomende data by dié in die sentrale woordelys te voeg nie (byvoorbeeld die opname as agterteks van geografiese name en hulle afleidings, natuurlik wat binne die beperking *S-Skooi* val).

'n Bykomende oorweging is dié van Gouws (2007) (sien ook Feinauer 2007: 264-265), naamlik om die agtertekste te gebruik vir gebruiksfunksies waarvoor die sentrale woordelys nie (primêr) voorsiening maak nie. As voorbeeld gee hy die gebruik van illustrasies en name van vleissnitte of 'n wêreldkaart met 'n verskeidenheid geografiese inligting om die kognitiewe funksie van 'n woordeboek te ondersteun (Gouws 2007: 84-85). In die beoordeling van die invoeging of weglating van agtertekste geld ook beperkings van uitgawes, personeeltekorte en voltooiingsdatums, maar veral ook 'n grondige besinning en empiriese navorsing oor die beoogde funksies van die WAT (sien die bespreking hierbo).

Die sentrale woordelys

Die sentrale woordelys vorm die hoofteks van elke deel van die WAT. Dit is egter dikwels ook die moeilikste aspek van 'n woordeboek om te beoordeel omdat 'n resensent meestal nie so 'n omvattende woordeboek van die eerste tot die laaste bladsy sistematies deurwerk en elke teks, artikel, databewerking, ensovoorts evalueer aan die hand van relevante kriteria nie. Soos Svensén (2009: 482-485) ook aanbeveel, moet die resensent volstaan met steekproewe wat beoordeel word teen die redaksionele beleid van die samestellers van die woordeboek self. Alhoewel Svensén (2009: 482) die beleid van die samestellers van 'n woordeboek as primêre evalueringskriterium voorhou, meld hy ook dat baie woordeboeke nie hierdie inligting verskaf nie. En soos hierbo aangevoer, word in spesifieke woordeboeke dikwels net vaagweg aan hierdie sake aandag gee. In sulke gevalle is die resensent dus, soos Svensén (2009: 482) aanvoer, afhanklik van algemeen aanvaarde evalueringskriteria wat onafhanklik bestaan van die woordeboek wat geresenseer word.

In resensies van die dele van die WAT, sien byvoorbeeld Feinauer (2007) en McLachlan (2013), word meestal al twee hierdie werkwyses gevolg. Aspekte van die sentrale woordelys wat meestal aan bod kom, is 'n kritiese evaluering van die redaksionele beleid van die WAT en die ooreenstemming met die beleid en die makro- en mikrostrukturele kenmerke van die sentrale woordelys (byvoorbeeld redaksionele inkonsekwentheid in die toepassing van hul eie beleid, soos weglating van lemmas of lemmatipes, die opname van moeilik interpreteerbare lemmas as onverklaarde lemmas, foute of gebreke in die bewerking van spesifieke lemmas, meer spesifiek, probleme met die bewerking van spesifieke datakategorieë, soos spelling, uitspraak, woordsoort, etimologie,

morfologie, leksikale definisies, verwysingsmetodes, gebrekkige gebruik van ortografiese struktuurmerkers met die onderskeiding van datatipes, ensovoorts.) Hierdie soort probleme is reeds goed bekend vir resensente en samestellers van die dele van die WAT en dit hoef nie verder ten opsigte van WAT Deel XIV uitgewys te word nie. Dié is sake waaraan die samestellers aandag moet gee (met hersienings en die beplanning en uitvoering van nuwe dele) en resensente hulle eie evaluering op goedgeмотiveerde kriteria moet baseer. Innoverend is wel die wyse waarop resensente voorstelle vir veranderinge en verbeteringe aan 'n deel van die WAT voorstel op basis van ontwikkelinge op die gebied van die leksikografieteorie en -praktyk. Sulke bydraes dien om leksikograwe hulle eie praktyk voortdurend krities te laat bekyk en, waar nodig, met innoverende oplossings vorendag te kom.

Uiteraard is resensies nie net gemik op kritiek nie. Vir sover woordeboeke konsekwent aan hulle redaksionele beleid hou, onder meer, word positiewe beoordeling van verskeie aspekte van die dele van woordeboeke gegee (sien byvoorbeeld weer Feinauer 2007 en McLachlan 2010). Wat die WAT en ook WAT Deel XIV betref, geld die omvattende opname van lemmas en hulle bewerking. Daar is geen gebrek aan voorbeelde hiervan as mens in hierdie deel van die WAT begin lees nie. Mens staan telkens opnuut verbyster oor die omvang van die Afrikaanse woordeskat wat daarin opgeneem is en die omvang van hulle bewerking.

Ten slotte

Een van die evalueringkriteria van goeie resensies is dat die slotevaluering van 'n woordeboek moet korreleer met voorafgaande evaluering van die ontwerp-aspekte daarvan (sien die bespreking hierbo). Die voorafgaande paragraaf oor die positiewe kenmerke van die WAT sou as 'n oortreding van hierdie beginsel gesien kan word. Hierdie artikel is egter nie uit en uit 'n resensie van WAT Deel XIV nie, maar eerder 'n artikel in die kategorie resensie-artikel vir sover dit 'n poging is om resensies oor die algemeen op 'n stewiger teoretiese voet te plaas, veral ten opsigte van evalueringkriteria. Soos die praktyk laat blyk, steur samestellers van woordeboeke hulle dikwels nie aan die evaluering van resensente nie, nog minder aan die jongste ontwikkelinge op die gebied van die leksikografieteorie en -praktyk. Dikwels gaan dit om ander, ewe geldige redes, maar as aanvaar word dat die leksikografiepraktyk positief deur die leksikografieteorie gelei moet word, dan is dit van belang dat mens weer krities oor die evalueringkriteria van leksikograwe en resensente moet besin.

Bronnelys

Woordeboeke

- Van Schalkwyk, D.J. (Hoofred.). 2000. *Woordeboek van die Afrikaanse Taal. Elfde Deel*. Stellenbosch: Buro van die WAT.

- Botha, W.F. (Hoofred.).** 2005. *Woordeboek van die Afrikaanse Taal. Twaalfde Deel*. Stellenbosch: Buro van die WAT.
- Botha, W.F. (Hoofred.).** 2009. *Woordeboek van die Afrikaanse Taal. Dertiende Deel*. Stellenbosch: Buro van die WAT.
- Botha, W.F. (Hoofred.).** 2013. *Woordeboek van die Afrikaanse Taal. Veertiende Deel*. Stellenbosch: Buro van die WAT.

Ander

- Atkins, B.T.S. en M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Bergenholtz, H. en T.J.D. Bothma.** 2011. Needs-adapted Data Presentation in e-Information Tools. *Lexikos* 21: 53-77.
- Boshoff, S.P.E.** 1926. 'n Standaardwoordeboek van Afrikaans. *Gedenkboek ter eere van die Genootskap van Regte Afrikaners (1875-1925)*: 307-328. Potchefstroom: Afrikaanse Studentebond.
- Botha, W.** 2004. Die impak van kritiek op die samestelling van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 14: 1-16.
- Botha, W.** 2006. Kom ons doen dit vir Afrikaans. Die WAT word 80. *Eikestadnuus*, 24 Maart 2006.
- Chan, A.Y.W. en A. Taylor.** 2001. Evaluating Learner Dictionaries: What the Reviews Say. *International Journal of Lexicography* 14(3): 163-180.
- Combrink, J.G.H.** 1962. 'n Prinsipiële beskouing oor WAT IV. *Tydskrif vir Geesteswetenskappe* 2(4): 199-221.
- Combrink, J.G.H.** 1979. Die sesde deel van die W.A.T. *Standpunte* 140, 32(2): 49-64.
- Cumming, G., S. Crop en R. Sussex.** 1994. On-line Lexical Resources for Language Learners: Assessment of Some Approaches to Word Definition. *System* 23(3): 369-377.
- Feinauer, I.** 1996. Die negende deel van die *Woordeboek van die Afrikaanse Taal*. *Lexikos* 6: 233-271.
- Feinauer, I.** 2007. Doen die *Woordeboek van die Afrikaanse Taal Deel XII* dit vir die Afrikaanse taal en die Suid-Afrikaanse leksikografie? *Lexikos* 17: 259-277.
- Fontenelle, T. (Red.).** 2008. *Practical Lexicography: A Reader*. Oxford: Oxford University Press.
- Gouws, R.H.** 1985. Die sewende deel van die *Woordeboek van die Afrikaanse Taal*. *Standpunte* 178, 38(4): 13-25.
- Gouws, R.H.** 1994. *Woordeboek van die Afrikaanse Taal, Deel IX*. Bespreking gelewer op die SAUK-radioprogram *Skrywers en Boeke*.
- Gouws, R.H.** 2007. A Transtextual Approach to Lexicographic Functions. *Lexikos* 17: 77-87.
- Gouws, R.H.** 2011. The *Woordeboek van die Afrikaanse Taal*. Grundlingh, A. en S. Huygen (Reds.). 2011. *Reshaping Remembrance. Critical Essays on Afrikaans Places of Memory*: 137-143. Amsterdam: Rozenberg Publishers.
- Gouws, R.H.** 2014. Leksikografie. Carstens, W.A.M. en N. Bosman (Reds.). 2014: *Kontemporêre Afrikaanse Taalkunde*: 373-407. Pretoria: Van Schaik.
- Gouws, R.H. en D.J. Prinsloo.** 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN MeDIA.
- Grobler, H.** 1978. 'n Voorlopige toepassing van S.P.E. Boshoff se kriteria vir 'n groot woordeboek op WAT I-VI. *Klasgids* 12(4): 29-46.
- Hausmann, F.J., O. Reichmann, H.E. Wiegand en L. Zgusta (Reds.).** 1989-1991. *Wörterbücher. Ein*

internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie. Berlin/New York: Walter de Gruyter.

- McLachlan, J.D.** 2010. Resensie. *Woordeboek van die Afrikaanse Taal, Dertiende Deel: R. Lexikos* 20: 750-756.
- Nielsen, S.** 2003. Anmeldelse af ordbøger. *LexicoNordica* 10: 27-44.
- Nielsen, S.** 2006. A Functional Approach to User Guides. *Dictionaries. Journal of the Dictionary Society of North America* 27: 1-20.
- Nielsen, S.** 2009. The Evaluation of the Outside Matter in Dictionary Reviews. *Lexikos* 19: 207-224.
- Odendal, F.F.** 1961a. Leksikografiese probleme I. *Standpunte* 14(6): 53-61.
- Odendal, F.F.** 1961b. Leksikografiese probleme II. *Standpunte* 15(1): 49-54.
- Odendal, F.F.** 1962. Leksikografiese probleme III. *Standpunte* 16(5): 48-55.
- Ripfel, M.** 1989. *Wörterbuchkritik. Eine empirische Analyse von Wörterbuchrezensionen.* Lexicographica Series Maior 29. Tübingen: Max Niemeyer.
- Svensén, B.** 2009. *A Handbook of Lexicography. The Theory and Practice of Dictionary-making.* Cambridge: Cambridge University Press.
- Swanepoel, P.** 2000. Providing Lexicographic Support for SL Vocabulary Acquisition: What Kind, under What Conditions, for Whom, and Why? Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000:* 403-418. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.
- Swanepoel, P.** 2008. Towards a Framework for the Description and Evaluation of Dictionary Evaluation Criteria. *Lexikos* 18: 207-231.
- Swanepoel, P.** 2013. Evaluation of Dictionaries. Gouws, R.H., U. Heid, W. Schweickard en H.E. Wiegand (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography:* 587-596. Berlin/New York: De Gruyter Mouton.
- Tarp, S.** 2004a. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.
- Tarp, S.** 2004b. Reflections on Dictionaries Designed to Assist Users with Text Production in a Foreign Language. *Lexikos* 14: 299-325.
- Tarp, S.** 2009. Reflections on Lexicographical User Research. *Lexikos* 19: 275-296.
- Van Sterkenburg, P. (Red.).** 2003. *A Practical Guide to Lexicography.* Amsterdam/Philadelphia: John Benjamins.
- Wiegand, H.E.** 1996. A Theory of Lexicographic Texts: An Overview. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(4): 134-149.
- Zgusta, L.** 1971. *Manual of Lexicography.* The Hague: Mouton.

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume I: 1567–1784.* 2004, xii + 259 pp. ISBN 0 19 925471 0 (Hb.). Oxford: Oxford University Press. Price: £115.

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: £115.

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume III: 1859–1936.* 2009, xxiv + 489 pp. ISBN 978 0 19 954937 5 (Hb.). Oxford: Oxford University Press. Price: £94.

Julie Coleman. *A History of Cant and Slang Dictionaries. Volume IV: 1937–1984.* 2010, xxvi + 503 pp. ISBN 978 0 19 956725 6 (Hb.). Oxford: Oxford University Press. Price: £85.

I

Now that four volumes of Julie Coleman's documentation of cant and slang dictionaries have appeared, it is a suitable time to consider them together* to determine the way she researched their history and the way she presented the results.

The volumes are structured almost similarly, with small variations made necessary by the special requirements of the different periods. Important in the front matter is, except the Prefaces, the Introductions in which Coleman presents, apart from a brief summary of the contents of each volume, the time delimitation of the dictionaries she discusses and the objectives she set herself.

She uses specific labels to distinguish the different types of non-standard language. Colloquial language is the language of conversation. Dialect terms are restricted to a geographical region. Colloquial and dialect terms remain peripheral in the dictionaries studied. Slang, which is usually short-lived, often belongs to specific age and social groups. Jargon is the specialized language of occupational or interest groups. Cant is the secret language of thieves and beggars, used for deception and concealment. Flash refers specifically to the fashionable slang of London's eighteenth- and nineteenth-century demi-monde. As these types of language cannot be clearly delineated, terms move easily between categories, being adopted by new groups of speakers. Not all lexicographers, whose dictionaries Coleman discusses use them with these meanings. Some lexicographers make little distinction between them, combining them all in one. Linking all these language types is their not belonging to Standard English. This determines the criteria for inclusion.

Coleman aimed at discussing these dictionaries within the context of the general and specialized dictionaries of a specific period, at demonstrating how the dictionaries are related to each other, at establishing their relationship with earlier glossaries, at identifying distinctive features of content and methodol-

* A review by Anne McDermott of the first two volumes appeared in *Lexikos* 17: 442-445, 2007.

ogy, and at placing them in the historical and social contexts in which they originated. In this way the dictionaries' topicality and role could be determined.

In each chapter related dictionaries are discussed, compared and evaluated. If available, short biographies of the compilers are given with reference to their place of birth, education, and sphere of activity. Features such as sources, subject matter, semantic coverage, usage labels and lexicographical features (citations, examples, authorities, etymologies, pronunciation, and cross-references) are considered, the author continually referring to the statistical tables of these features given in the Appendixes of each volume. These statistical tables therefore form an integral part of the discussion of the different dictionaries in the main text of a particular volume. Each chapter is rounded off by a summary of and conclusions about its contents. The concluding chapters give a general overview of the trends observed in the different publications of the period covered in a specific volume. The back matter of each volume contains, apart from the Appendixes, an extensive Bibliography, a Subject Index and a Word Index.

In the following four paragraphs a short survey of the contents of the four volumes is given, attempting to indicate the breadth and depth of Coleman's research.

II

Volume I covers slang dictionaries published between the years 1567 and 1784. These lists must be understood against their historical background, which Coleman describes in Chapter 1. The early part of the period is characterized by "the agrarian revolution, and the progressive criminalization of poverty" and the latter part by "the transition to a cash culture and the haphazardly harsh treatment of criminals" (Volume I, p. 8). The lists reflect "changing interests and concerns caused by population growth and migration, developments in legal and penal theory and practice, and far-reaching changes in trade and travel" (Volume I, p. 8).

Chapter 2 deals with Thomas Harman's list, starting off with the short glossary of beggar's language in *Caveat or Warening for Common Cursetors* (1567). Harman's list was reworked and republished by Thomas Dekker in his *Bellman of London* series (1608), while S.R. published a list in *Martin Mark-all, Beadle of Bridewell* (1610) as a riposte to Dekker's first book.

Chapter 3 discusses the three cant lists produced by Richard Head in *The English Rogue* (1665), each based on Harman's list, but every time with something new added. These were followed by Head's *Canting Academy* (1673) and its different versions. Head's list was adapted and reissued until the end of the eighteenth century. It was also included in editions of *The Triumph of Wit* (1688) by John Shirley. Editions published in 1707, and later, contain a number of deviant readings not found in earlier ones. The cant list *The Life and Death of the English Rogue* (1720) published many years after Head's decease, is derived

from that in the *Canting Academy*. It shares some readings with a list in *The History and Actions of Jonathan Wild* (1725) which appeared in two later editions.

B.E.'s *New Dictionary of the Terms Ancient and Modern of the Canting Crew*, published in 1698 or 1699, is analysed in Chapter 4, being the most substantial work which appeared during the period covered by Volume I. Coleman mentions that B.E.'s dictionary represents a major development in the English slang and cant dictionary tradition for at least four respects: it is the first word-list published in an independent volume, the word-list is expanded to over 4,000 entries compared with Head's list of 265 in the *Canting Academy*, the new entries represent an expansion in the scope of the contents, including, apart from cant, also general slang, colloquial language and jargon, and it contains some of the features commonly found in mainstream monolingual and bilingual dictionaries. B.E.'s work was quickly and comprehensively pirated. One group of word-lists by Carew, derived from the anonymous *New Canting Dictionary* (1725), was produced in such large numbers that Coleman treats them in a separate chapter, Chapter 5. The later lexicographers, who based their dictionaries on B.E.'s each ended up with a different selection of terms, showing the breadth of its coverage. Only after several editions of Alexander Smith's popular work *The History of the Lives of the Most Noted Highway-men* (1714) had appeared, did he include a cant glossary in the augmented, three-volume edition of the 1719 *A Compleat History of the Lives and Robberies of the most Notorious Highwaymen*. The editor of the American edition, which only appeared in 1813 largely excluded the cant terms, producing a less offensive, more moralistic dictionary of slang. B.E. made considerable contributions to the history of English in general and its slang in particular. Coleman mentions that, through Grose's dictionary, many of B.E.'s entries appear in eighteenth- and nineteenth-century word-lists, still presented as contemporary slang.

Chapter 5 deals with the cant glossaries contained in the editions of Bampfylde-Moore's life, the first edition published in 1745, and the first to include the glossary, appearing in 1750. His biography was supplemented, abridged, edited and adapted for more than a century under various titles which reflect the changes of taste during the years. The word-lists are a mixture of plagiarism from various sources and original observation of changes in slang. They are not, however, as claimed, glossaries of the language of English Gypsies.

In Chapter 6 Coleman discusses cant and slang lists not wholly derived from the four main strands of the tradition: Harman, Head, B.E. and Carew. A varied selection, they are attached to works as different as exposés of the vices of the capital, memories and confessions of real-life and fictional villains, miscellaneous anthologies, and plays. After Robert Greene's cony-catching works, published during 1591–1592, no wholly independent cant glossaries were produced for almost a century. Between 1688 and 1799 there appeared several of these lists which are frequently authenticated by their compilers' assertion of inside knowledge. Greene associated with dubious characters in London, while biographical evidence supports Charles Hitchen, James Dalton, Daniel Defoe

and John Poulter's position of authorities on criminal and prison life and language. These lists might contain convincing evidence of contemporary cant: *The Regulator* (1718) by Hitchin, *Genuine Narrative of all the Street Robberies Committed since October last* (1728) by Dalton, *Street Robberies Considered* (1728) by Defoe and *The Discoveries* (1753) by Poulter.

Although numerous general dictionaries included at least some cant terms, especially two of those Coleman discusses in Chapter 7 are significant to the cant and slang dictionary tradition. Elisha Coles not only incorporated some cant in his *English Dictionary* (1676), but also influenced B.E.'s dictionary. Nathan Bailey not only provided a separate list of cant terms in *The Universal Etymological English Dictionary* (1737), but it was also used by Grose in the compilation of his dictionary.

In the concluding chapter, Chapter 8, Coleman mentions that only a few cant and slang lists produced during the period covered by Volume I were sold as independent books: most formed only a small part of some larger work. The sixteenth-century glossaries are typically found in rogue books, which are general descriptive catalogues of the various types of villain that the reader should learn to avoid. By the seventeenth century, the word-lists normally form part of the biographies or autobiographies of specific criminals. The only exception is B.E.'s dictionary, which was directed wholly to public interest in non-standard language. In spite of its wider linguistic focus, those who used his dictionary as the basis for their own, mostly concentrated on his cant terms.

These cant and slang word-lists and the works in which they were found, all depended on and contributed to a general fear of crime. These books dealing with criminal lives fulfilled a deterrent function, relying on indoctrinating and admonitory example. Coleman believes that a wide range of readers would have bought these books: those seeking entertainment, an understanding of criminality and advice for the safeguarding of their lives and possessions.

There is no conclusive evidence that the authors of these works had any first-hand knowledge of the criminal underworld. Although the contents of these lists are often similar, they claim to cover the language of a wide range of users. However, the fieldwork methods of these compilers should be viewed critically and the results with reservations. Coleman stresses that these word-lists "not only form part of lexicographical tradition, they are also intricately linked with contemporary literary depictions of the criminal underworld" (Volume I, p. 187). Even if the contents of these lists, according to Coleman, were terms used by criminals, they may represent little more than short-lived code words employed within specific criminal gangs. The discarded vocabulary, if not fallen into disuse then, continued to be freely included in word-lists and used in literary works as "convenient symbols of criminality" as Coleman puts it (Volume I, p. 189). Widely known in the eighteenth century, canting drama and ballads, examples of which are given in Appendix B, were composed by hack writers and street singers themselves, rather than by criminals. In the word-lists and literary works, there are undoubtedly some terms that are

genuine criminal cant, but by their repeated occurrences, became fossilized. Other terms which acquired more general currency, may have formed part of contemporary urban slang or dialect, while a few are still in common use.

III

Volume II considers cant and slang dictionaries published between 1785 and 1858. During this period, language was felt to be changing as rapidly as society. The working classes were abandoning rural pursuits and dialects in their migration to the industrial towns and cities. Opportunities for discussing social and linguistic changes were provided by numerous clubs and periodicals. The circumstances were suitable for the publication of an authoritative lexicon of slang.

The three editions of Francis Grose's *Classical Dictionary of the Vulgar Tongue* (1758, 1788, 1796), covering not only general slang, but also cant, jargon and dialect, are analysed and compared in Chapter 1. Trying to appear authoritative and scholarly, Grose listed his sources both in the preface and in the body of the dictionary, making them seem more extensive than they actually were. In the first edition, he chose from his dictionary sources those labelled as 'cant', those illustrated by citations and/or those dealing with selected semantic areas, thereby including relatively few terms not listed earlier. Grose's other sources, credited where appropriate, comprise canting works, novels, plays, popular literature and reference works. The second edition shows a considerable augmentation of the first. In addition to the use of new antiquarian sources, he again consulted volumes used in the compilation of the first edition. By including more general non-standard and colloquial language, he moves away from the language of criminals. The third edition, extensively edited by Wigstead, may include some material from notes made by Grose. Although it contains relatively few new entries, many existing entries were rewritten and updated. With an increase in jargon, the movement away from cant continued.

Grose's dictionary established his reputation as an important authority on both antiquarian and contemporary slang. Later lexicographers relied on this reputation by using either his name or his work. The dictionaries discussed in Chapter 2 range from the entirely pirated to the partially original. The most important dictionary is the *Lexicon Balatronicum* (1811) which increased the word-list by over 300 headwords and almost 400 new entries. The compiler edited existing entries to make them more concise, correct, modern and informative, adding citations and authorities. George Matsell's *Vocabulum* (1859), although largely derivative, made, through his additions, a significant contribution towards charting early American slang, demonstrating an awareness for small shifts in meaning. Herbert Astbury's selection from Matsell's dictionary in *Gangs of New York* (1927) resulted in a concise, more sensational glossary.

Although the dictionaries discussed in Chapter 3 are all ultimately derived

from that of Grose, they are linked by a common intermediary, Humphrey Tristram Potter's *New Dictionary of all the Cant and Flash Languages* (1795). Over 80 per cent of the word-list of Potter's dictionary is found in earlier slang dictionaries from which he selected terms labelled as 'cant' and dealing with crime and punishment. By deleting citations, authorities and etymologies given in his sources, Potter produced a more concise dictionary. George Andrewes's *Dictionary of the Slang and Cant Languages* (1809) is largely derived from Potter's with the introduction of a few terms of his own. The tiny anonymous *Flash Dictionary* (1821) edits the material that it adopts from Andrewes and Potter, adding entries which represent contemporary cant and slang. George Kent's definitions in *Modern Flash Dictionary* (1835) are usually briefer than those of his sources. Just over a fifth of his entries cannot be traced to these sources, thus indicating that these probably reflect contemporary speech. The author of *The Sinks of London* (1848) adopted his entire word-list from Kent's. In comparison with other dictionaries in this group, Duncombe's *New and Improved Flash Dictionary* (c. 1850) added significantly to the material from its sources, especially Kent's dictionary on which it is based. Although only three-quarters of the size of Kent's dictionary, over two-fifths of its entries were entirely new. The dictionaries related to Potter scarcely testify of original lexicography, mostly only adding a few new entries to those taken from their sources.

Chapter 4 deals with the word-list in *The Memoirs of James Hardy Vaux* (1819). Coleman is of the opinion that this glossary is a genuine record of the language used by English felons at the beginning of the nineteenth century. It demonstrates the dissemination of English slang and cant across the world, confirming the continued currency of many terms found in the *Lexicon* and Potter's dictionary.

Known for the journal *Boxiana* (1818–1824) as well as the book *Life in London* (1821), Pierce Egan brought out in 1823 an edition of Grose's dictionary, which is discussed in Chapter 5. As sources he used several minor slang dictionaries, in addition to the *Lexicon* and Vaux's word-list. In selecting terms from the *Lexicon*, he tended to omit marginal terms, concentrating on its core semantic coverage. He also re-alphabetized the word-list.

The topic of Chapter 6 is the work of John Badcock, who published works on boxing and racing under the pseudonyms John Bee and John Hind between 1816 and 1830. Each of Egan's most popular works was preceded by a similar but less successful work by Bee: *Boxiana* by Bee's *Lives of the Boxers*, and *Life in London* by Bee's *Letters from London*. Egan also published a slang dictionary before Bee could finish his. Throughout his dictionary Bee tries to demonstrate his greater knowledge and erudition. He provides anecdotal and encyclopaedic information, moral and social commentary and guidance on usage. He abandons crime and punishment for social and sporting terms. Bee claims that he built on Grose's dictionary. Like Grose he draws attention to his comprehensive written sources which he, unlike Grose, used critically. Both attempted, by means of the extensive use of citations and authorities, and particularly in Bee's

case, etymologies, to produce scholarly and authoritative reference works. Both emphasized their use of slang-speaking informants, Bee, however, commenting on non-standard pronunciation more than Grose. Bee even provided some headwords with grammatical labelling. Coleman is of the opinion that Bee's dictionary is in many ways better than either the *Lexicon* or Egan's dictionary. Its coverage is more limited both as far as etymology and register are concerned, thereby showing that Bee had a clearer and more focused purpose. Egan's dictionary was largely derivative, while Bee recorded contemporary slang, especially sport slang, accurately and observantly.

In Chapter 7 Coleman treats the numerous minor British cant and slang word-lists that appeared during the late eighteenth and early nineteenth centuries. These were not issued as dictionaries in their own right, but formed part of larger works such as plays, autobiographies and guidebooks, being added as appendixes or glossaries to these. These shorter lists were meant for a public not able to afford the more extensive volumes produced by compilers like Grose, Egan and Bee. Independent of the main slang dictionary tradition, these lists contributed significantly to the recording of non-standard English during this period. Their authors include not only journalists and hack writers, but also criminals and officials tracking down crime, both of whom should be expected to be knowledgeable of the secret languages used by contemporary villains. Several of these glossaries profited by the craze for flash language arising from the success of Egan's *Life in London*.

In contrast to Noah Webster who created an authoritative dictionary for Americans, John Bartlett restricted his dictionary to colloquial, dialectal and slang terms peculiar to the United States, but excluded cant. The American cant and slang glossaries discussed in Chapter 8 are independent products by users or hearers of the terms they include. Unlike Webster's more general and Bartlett's more specific dictionaries, they rely rather on spoken than written sources, being therefore useful portrayals of the development of American cant and slang.

This volume is concluded by Chapter 10 in which Coleman indicates, in addition to the continuities in slang lexicography, the changes in linguistic theory and practice, in attitudes towards non-standard registers and in publishing. As far as continuities are concerned, two examples can be mentioned: some lists, like Matsell's, compiled first-hand, can trace their ancestry, and even some of their contents, to Harman's *Caveat*; and some lists, like those of Grose and Egan, compiled, at least in part, from written sources at their disposal, copied B.E.'s example.

This period, however, saw the development of a historical approach to slang lexicography as practised by Grose. Because of Grose's influence, at first directly and later indirectly through Egan, many dictionaries included the same core word-list. Shorter independently produced lists often better reflected current non-standard language than the more generally consulted works like the *Lexicon*. By changing into flash, cant has now become stylish. While it still

implied knowledge of the underworld, the use of flash did not indicate criminality in a similar way as the use of cant. What made flash appealing was, according to Coleman, that "it occupied the area between the classes: between those who were respectable and those who were not" (Volume II, p. 260). Flash language was therefore, for users of any class, an expression of defiance. Publishers were responding to market trends. After the appearance of Egan's *Life in London*, for example, the inclusion of a slang glossary could improve the sales of a similar work. These works were bought mostly for entertainment. Their readers could raise their status among their peers by using the same flash language in everyday life.

IV

Volume III deals with cant and slang dictionaries published between 1859 and 1936. In the late 1850s, the British Empire was at its height. By the end of the nineteenth century, however, the Empire was declining. At the beginning of the First World War, America was both more industrialized and more productive than the United Kingdom. As America's national influence and confidence continued to grow, American English disengaged itself from British English, accentuating its greater informality and creativity. English in Australia took longer to assert its independence.

The different editions of John Camden Hotten's *Dictionary of Modern Slang, Cant and Vulgar Words* (1859) are discussed in Chapter 1. Hotten tried to do for the nineteenth century what Grose had done for the eighteenth: to summarize the history of cant and slang, and to provide an account of contemporary non-standard usage. It catered for an uncritical audience and a popular market. Although Hotten emphasized his use of slang-speaking informants, he also extensively employed numerous written sources, praising some and condemning others, in this way appearing to produce a scholarly and authoritative reference work.

The lexicographers of the international slang dictionaries treated in Chapter 2 were not native speakers of British English. Their dictionaries often include information about slang in other European languages. Apart from Henry Baumann's dictionary, those of Albert Barrère, Charles Leland and Karl Lentzner are discussed. These works are of a scholarly nature, beyond the understanding and means of the uneducated.

Of the works considered in Chapter 3, the most important is John Farmer and William Henley's *Slang and its Analogues, Past and Present* (1890–1904). Bringing about many improvements in English slang lexicography and having a great influence on later slang lexicographers, Farmer and Henley's dictionary gave fully referenced citations both to support their definitions and to indicate dates of usage. Speculation about unknowable etymologies was avoided. Obscene and offensive terms were defined without being evasive or obscure. They developed a stricter sense of the boundaries of slang, excluding dialectal

and colloquial terms.

As shown in the discussions of the British general slang dictionaries in Chapter 4, slang lexicography conducted in Britain during this period was predominantly historical. *A Dictionary of English Slang and Colloquialisms* (1913) by Arthur H. Dawson and *Slang Today and Yesterday* (1933) by Eric Partridge are the only glossaries offering even a limited insider-perspective and that only in the relatively few original entries. Looking back to the nineteenth from the early years of the twentieth century and the inter-war period, the discontinuity between the past and the present became apparent. American English and its slang were now distinct enough from their British equivalents that mutual explanation was required, glossaries of American slang for British audiences being the more common. Rhyming slang, indicating a particular image of the British urban working classes and capturing the imagination of those hearing it, both at home and abroad, caused a disproportionate number of these dictionaries to appear among dictionaries of British slang.

Chapter 5 covers glossaries of British school and university slang, sometimes appended to publications such as guides and reminiscences. Many schools developed their own slang, some persistent enough to be documented. Such slang remains an important part of the identities of some schools, several providing a glossary for the information of new pupils. The enforcement of conformity and the instilment of respect for school traditions were important functions of school slang. These glossaries are generally reminiscent, associated with memoirs of school life in earlier days to be seen in light of a general pressure to modernize and reform.

The slang dictionaries discussed in Chapter 6 are accounts of English used in Australia, largely by British emigrants. Given the tendency of glossaries of Australian slang to include widely accepted Australianisms, there is sometimes little difference between the Australian slang dictionaries and the general dictionaries of Australian English. Coleman mentions that slang is a particularly problematic label for Australian English, because within the inherent informality of Australian English many slang terms are acceptable, and because many distinctly Australian words and phrases were stigmatized as slang in the nineteenth century. The use of English was largely viewed from the perspective of British English, resulting in deviations from Standard British English to be regarded as merely uncouth language.

The American slang dictionaries considered in Chapter 7 are characterized by their diversity: slang could not become national without the power of the media. At the beginning of this period no contacts and therefore no shared interests could develop across a vast continent. More possibilities for personal encounters with American English were supplemented and eventually overshadowed by the recorded versions distributed by film and record companies. Several of these glossaries were produced for British theatre- and cinema-goers who found American terms unfamiliar and unintelligible. The glossaries for an international audience focus on similarities rather than differences between

speakers of American English, probably therefore contributing to the development of the sense of an existing national slang.

Chapter 8 concentrates on American school and university slang dictionaries. In some respects these contain the same kind of insider slang that characterizes British school and college glossaries. All the glossaries discussed in this chapter date from the later nineteenth century when the enrolment of students increased. These glossaries are untrustworthy in methodology, terminology and documentation: their authors usually do not inform on the way the lists were collected, define what kind of slang are being listed and give little particulars about currency. They document the beginnings of a youth culture that was to move beyond educational affiliation. Here, however, Coleman stresses, slang was often a measure of conformity, rather than rebellion. These glossaries are better than contemporary British school and college lists: they contain current slang collected by young academics, and are therefore more reliable than those recorded in older men's reminiscences.

Chapter 9 examines dictionaries of the slang of World War I. Because of the Americans' shorter involvement in the War and because there was already an established vocabulary at their arrival, relatively few of these glossaries are American. Some military slang was carried over from the pre-War period into the First World War. Many of the glossaries comment on differences in slang used in various areas and periods of the War. After the War, some terms survived into wide or restricted usage, but the ones becoming obsolescent served often as motivation to document them. These glossaries, having many different purposes, cannot be interpreted without reference to their function. They reveal details not only of the military experience, but also of personal and social interaction. The earliest of these glossaries appeared while the War was still in progress; the latest were published when the next war had become inevitable.

The effects of World War I on American society and language are shown in Chapters 10 and 11, which deal with glossaries of tramp and criminal language. There is some continuity between the language of tramps and criminals, and, as Coleman says, some glossaries could justifiably have been placed in either chapter. In contrast with the period covered in Volume II, there are relatively few British tramp and criminal glossaries.

Most of the glossaries of homelessness in Chapter 10, dealing with the language of tramps and hoboes, are American. There is a sufficient overlap between the numerous unrelated lists to suggest that there really was some national link joining the terms. These glossaries demonstrate the appeal of an itinerant lifestyle comparable with the fascination with Gypsy life found in eighteenth-century British texts.

Few of the dictionaries of crime discussed in Chapter 11 are British and Australian. American cant glossaries tend to categorize criminals and criminal activities into specific groups, documenting and labelling the language of these groupings separately. These lists often lacked authenticity which was counteracted in the late 1920s by sociological studies of criminals and their language.

David W. Maurer changed cant lexicography by introducing data collection in the field, his presentation of separate lists for different criminal communities emphasizing the specialization of crime. In the late 1920s particularly, American criminals started to reveal their secret language as part of the process of reformation. By the 1930s, glossaries tended to concentrate on the language of prisons rather than that of unjailed criminals.

Many of the glossaries of the slang of the entertainment industries, grouped together in Chapter 12, could also have been treated in Chapters 10 and 11. The entertainment industries were beginning to have an effect on the development and dissemination of slang. Academics in America and England needed to help the audiences understand the slang used in the respective countries. The entertainment industries also developed slang of their own, this chapter therefore concentrating on glossaries of terms used within show business in its various forms. The theatre, particularly vaudeville, the circus and the carnival were all on the decline after the early years of the twentieth century, the itinerant employees often living, as Coleman puts it, "on the edges of poverty and legality" (Volume III, p. 7). Like the glossaries discussed in Chapters 5 and 10, these glossaries were sometimes self-consciously preserving the vocabulary of a fast disappearing way of life. Similar to the glossaries treated in Chapters 10 and 11, they are mostly American. The later dictionaries discussed in Chapter 12 show the appearance of African-American music and language, laying the foundations for the rapid changes in musical trends after the World War II period.

The concluding chapter, Chapter 13, summarizes the general trends of slang lexicography during this period. Many of the British slang dictionaries, being historical, detail changes terms underwent through time or document obsolescent terms from previous generations. As social structures in Britain, the United States and Australia differed, each system of non-standard speech carried a different meaning. Although social and educational distinctions in language were maintained in the United States, contemporary slang and dialect were of interest for their own sake. While still considered inferior by Anglo-centrics, Australian slang and Australian English had become symbols of national pride and self-respect to their speakers. It is uncertain how reliable slang glossaries from this period are as representative of non-standard speech, but the areas of slang they chose to document indicate a contemporary interest. This period sees the necessity for and beginning of sociological slang lexicography, sometimes scholarly, sometimes journalistic.

V

Although there were still general dictionaries of national slang in the period 1937–1984 covered by Volume IV, with Partridge's dictionary as dominant example, lexicographers increasingly specialized in the slang of social sub-groups. These slang dictionaries, mostly by amateur lexicographers, reveal

their incompetence as historians of the language, etymologists, definers, grammarians and proof-readers. In this volume, as Coleman says, "glossaries that warn of the dangers posed to society by slang users are in the minority; instead authority and traditional values threaten individual self-determination" (Volume IV, p. 2). She largely organized the dictionaries geographically or thematically, with each group of dictionaries discussed chronologically.

Chapter 1 deals with the eight editions of Eric Partridge's *Dictionary of Slang and Unconventional English* in its four redactions. Partridge was not a discriminating lexicographer: his dates are often based on deduction rather than evidence, he frequently includes unreliable etymologies and he depended on single sources or correspondents for many of the terms listed. After the third edition no up-to-date coverage of contemporary slang was provided. In light of political and cultural influences, his exclusion of American slang became more difficult to justify. The separate listing of acronyms was hard to sustain and his complex system resulted in mistakes in alphabetization. Until Paul Beale's combining of the two lists and revision of the alphabetization, users were never certain whether a sought term was included.

Chapter 2 concentrates on dictionaries of slang used during periods of conscription. These treat glossaries of military slang of the Second World War, as well as glossaries of the First World War published during this period. Only two years after the start of World War I the first slang glossaries appeared, but already early in World War II their potential contribution to morale was recognized. The official American glossaries imply that all personnel were united by their shared slang, but in reality technological and tactical specialization led to the documentation of slang restricted to specific units and divisions. Military slang was incorporated into general slang during the War and for some time afterwards. This is especially the case with air force slang where civil aviation kept original military slang in general circulation. The long continuation of World War I slang lexicography was prompted, according to Coleman, by a feeling of futility, while the justness of World War II stood confirmed. Few glossaries of the Korean and Vietnamese wars focus on slang, perhaps because of the short period of conscription and the maintenance of civilian slang.

Chapter 3 discusses British slang dictionaries of the post-war period, tending to be historical in content or approach, which is particularly evident in their concentration on rhyming slang. These rhyming slang dictionaries, found over-represented among dictionaries of this period, are often associated with the entertainment business, emphasizing the unreality of the Cockney character presented in them. The relatively few number of general British slang dictionaries during this period must be ascribed either to Partridge's domination of slang lexicography or the overwhelming influence of American slang.

The declining influence of British English is reflected in Chapter 4 which groups together glossaries of the slang of new English-speaking nations, of which Australia was by far the most productive. As Australian English interacted with a developing sense of nationalism, its slang tends to be presented as

uniform and unifying. Lexicographers in New Zealand and Canada had to distinguish distinctive national forms not found in the usage of their neighbours Australia and America respectively. The production of national dictionaries in New Zealand and Canada during this period probably explains why less attention was paid to slang. Of the British colonies, only Nigeria produced slang dictionaries, indicating that national developments in non-standard English had taken place there.

As shown in Chapter 5, considerably more lists of American than British slang appeared during this period. This chapter also looks at several American glossaries dealing, often nostalgically, not only with circus and carnival slang, but also with soda-fountain and hotel slang. Coleman believes that patriotic motives may explain some of the academic attention paid to American dialect and slang. Although the formal English of Britain and America are not markedly different during this period, their slang is. Generally the American glossaries are more contemporary in focus than the British ones.

Chapters 6 to 12 focus on slang glossaries of social sub-groups, although grouping occurs nationally within these chapters. British school and college dictionaries and their American counterparts discussed in Chapter 6 represent the continuation of established traditions of school and college lexicography, motivated, at least partially, by the desire to distinguish one educational establishment from another by accentuating its unique traditions and language. However, by the late 1960s, as seen in Chapter 8, there is a move to study youth slang, rather than school and college slang. Coleman stresses that "it would ... be impossible to discuss youth slang during this period without reference to African-American slang" (Volume IV, p. 3). Throughout Chapters 7 and 8, she presents the assimilation of youth culture. Chapter 7 discusses glossaries of African-American slang, followed by glossaries of musicians' slang reflecting the way of speaking which accompanied each new musical trend. Chapter 8 deals with glossaries of slang used by counter-cultural groups and youth culture groups. These glossaries served a variety of different purposes, for example to challenge or defend traditional values, to help young people deal with the youth scene and to define or address a target market. Coleman treats them chronologically to emphasize the rapid dissemination and appropriation of the language of these groups by advertisers and the media. Philosophies of self-determination, social revolution and sexual liberation characterized youth culture during this period. These are also crucial to many of the glossaries of sexual orientation which Coleman examines in Chapter 9.

In contrast to the general acceptance of other slang users, two groups are considered a threat to society: drug users and criminals. Criminals were always seen as dangerous, but drug users became prominent during this period. Terms for drugs are found in glossaries discussed in most chapters in Volume IV, particularly frequently in Chapters 6 to 9, but Chapter 10 especially concentrates on glossaries focusing on the slang of drug addicts or dealers. Many of these glossaries were compiled by or for professionals working with addicts in

various capacities. The many glossaries of drug slang published during this period, especially in America, could, according to Coleman, be the result of the increasing problem of drug addiction in the inner cities, perhaps attributable to the association of drugs with successive popular music trends, or a result of the criminalization of drugs which increased the potential profits of importers or purveyors. Chapter 11 considers dictionaries of the language of other criminal groups, particularly prisoners. The adoption of American slang by criminals in other parts of the world should be attributed to the media rather than personal contact. The production of so many glossaries by prison officials and sociologists indicates that prison slang was successful in allowing prisoners to communicate secretly. It strengthens the belief that neither prison authorities nor prisoners themselves can be fully understood without knowledge of their language.

A few glossaries react against the liberal movements considered in Chapters 6 to 10 and the Citizens' Band Radio slang glossaries discussed in Chapter 12 reflect the call for a return to traditional values. These dictionaries and glossaries show the effect of changing technology on the development and transmission of non-standard terms, express a belief that first-hand knowledge qualifies a person as slang lexicographer and demonstrate that there is a market for accessibly written populist dictionaries. Coleman sees CB users as representing an early form of virtual community, with their lexicography being joined by a dictionary of the slang of computer users.

Conclusions about slang lexicography during this period are drawn in Chapter 13. Scholars discussed the processes and purposes of slang lexicography, some emphasizing the importance of sociological context, others the necessity of historical textual research. This also becomes apparent in many scholarly glossaries, which belong either to one or the other approach. Although slang dictionaries, largely composed by free-lance compilers, are unreliable for academic purposes, these lexicographers are not deterred by debates about the way of undertaking slang lexicography or delineating a particular area of interest. Being knowledgeable about slang, they compiled dictionaries, which sometimes came to be used as tools in the struggle for personal or national identity, civil rights, non-conformity, self-expression, sexual liberation, and a greater understanding and equalization of various social sub-groups.

VI

After studying these four volumes, the reader is left with two predominant impressions: first, the enormous amount and diversity of materials collected, consulted, compared and evaluated, and second, the competent integration and presentation of these materials in an absorbing and scholarly way. It is a masterly and admirable accomplishment. These volumes not only give an overview of the dictionaries that were published during particular periods, but also of the social and historical backgrounds which produced them. They are therefore

not only informative to lexicographers and linguists, but also to sociologists and historians. Hopefully in a few years' time, a fifth volume will appear that will cover the period from 1985 onwards.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal
Stellenbosch
South Africa
(jcmddp@sun.ac.za)

G.-M. de Schryver et al. (Eds.). *Oxford Bilingual School Dictionary: IsiXhosa and English.* 2014, 562 pp. ISBN 978-0-19-576682-0. Cape Town: Oxford University Press Southern Africa. Price R129.95.

Oxford University Press Southern Africa has added the *Oxford Bilingual School Dictionary: IsiXhosa and English* to its impressively developing series of bilingual school dictionaries involving English and other official languages of South Africa. The series already has similar dictionaries in other languages namely isiZulu, Northern Sesotho and Afrikaans and is hoped to include the rest of the official languages of South Africa. The dictionary under review is expected to be useful for school learners, educators and the general isiXhosa-speaking community which had to make do with old and outdated dictionaries, most of which were not specifically produced with school learners in mind, especially in the present South African linguistic and educational milieu. For African metalexigraphers, it is yet another moment of reflection on the direction that *school lexicography* is taking, and hopefully producing a symbiotic relationship between metalexigraphy and practical lexicography, given that the latter has been lacking with respect to school dictionaries.

Besides being produced by the global lexicographical powerhouse Oxford University Press, this dictionary is a result of collaboration by accomplished experts in their respective fields. Among others, Gilles-Maurice de Schryver is not only an experienced practical lexicographer, but also a globally-renowned metalexigrapher; Mary Reynolds and Daphne Paizee are popular names in English school dictionaries; Sydney Zotwana is a respected scholar and writer in isiXhosa; and Fikiswa Magqashela's name would be familiar to keen readers of isiXhosa literary works. Clearly, such a team should be expected to produce a school dictionary that can contribute immensely to language learning that involves both isiXhosa and English, as well as the development of isiXhosa into a fully functional official language in South Africa.

Although the target users of the dictionary are not clearly spelt out in terms of their school and language learning levels, as is the case in the other dictionaries in the series, it is clearly indicated that this dictionary "has been designed to help people learning isiXhosa to understand and use isiXhosa better; and to help people learning English to understand and use English better" (p. vi). However, in comparison with those dictionaries in the Oxford First Bilingual Dictionary Series whose target users are clearly foundation phase learners, the size of the *Oxford Bilingual School Dictionary: IsiXhosa and English* makes it obvious that it is targeted at the intermediate and senior school phases. It may be claimed that this dictionary is intended to build on the dictionary and language skills that are developed by the *Oxford First Bilingual Dictionary: IsiXhosa and English*.

As it intends to help the learners of both languages, this dictionary is therefore a bidirectional bilingual dictionary with two wordlists. The first wordlist provides lexicographical description of isiXhosa lemmata, while English

lemmata are dealt with in the second wordlist. We are told that approximately 5 000 words (pp. ix; xi), mostly drawn from corpora of the respective languages, are treated in each wordlist, whereby frequency of occurrence in each corpus is indicated for the top 1 500 words. Stars are used as frequency indicators, with three stars for the top 500, two stars for the next 500 and one star for the third 500 words. However, not only the highly frequent words are included. Words that are not even in the corpora are said to have been included on the basis of their importance in the educational situation of the learners (pp. ix; xi). From this, it may be accepted that beyond technological lexicographic applications, this dictionary was conceived out of a careful consideration of the user needs and user situations.

In the isiXhosa wordlist, besides frequency indication where appropriate, lemmata are described by means of word-class/type of speech data, explanations of meaning in English or translation equivalents and illustrative examples in isiXhosa and their English translations. In lexicographical terms, these types of data and their indicators constitute the obligatory microstructure (Gouws and Prinsloo 2005) of the isiXhosa-English part of the dictionary. However, there are other additional types of data and data indicators that apply only to specific lemmata. For example, noun class prefixes are indicated for lemmata which fall under the speech categories of nouns, possessives, adjectives and relatives together with the indication of the noun number (i.e. whether the noun is in singular or plural), providing information that does not apply to other types of lemmata. On the other hand, some lemmata are not provided with comprehensive treatment as they are clearly considered grammatical forms of others which are described in a more comprehensive manner in their appropriate alphabetical access positions. Relations between such and other lemmata are indicated by mediostuctural guidance in the form of an arrow and the guide word *See*. That way, the user is advised to seek more comprehensive treatment in the cross-reference address, thereby saving precious dictionary space while re-establishing lexical relations of related lemmata.

Another important aspect regarding the isiXhosa wordlist is the lemmatisation approach. Lemma selection and ordering of lemmata in Nguni dictionaries is an old debate that gets re-opened whenever a new dictionary is compiled or critiqued (cf. De Schryver 2008; 2010, Gouws and Prinsloo 2005; 2005a, Kropf 1899). Inevitably, to "make this dictionary exceptionally easy to use" (p. x) entailed adopting an approach that would ensure that the users of the *Oxford Bilingual School Dictionary: IsiXhosa and English* are able to find answers regarding words that prompt dictionary consultation and do so as punctually as possible. To this end, the compilers of this dictionary depart from the traditional method of listing nouns in isiXhosa and other Nguni dictionaries using the first letter of the stem, the so-called stem method, to the new method that lists nouns and other words using their first letters, the so-called word method. This sees words *abafundi* (learners), *ekufundiseni* (in the teaching), *isifundo* (lesson) and *umfundi* (learners) respectively lemmatised under letters **Aa**, **Ee**, **Ii** and **Uu**

respectively instead of letter **Ff** as would have been the case if the stem method had been used. This means that the dictionary users do not have to worry about dissecting the word into its different formatives to isolate the stem, which could be difficult for them and even insurmountable in some words. All they need to do is to look up the full word that they have encountered using its first letter. They will only look under letter **Ff** for information regarding the verb *-funda* (learn) and its included extended verbal forms. What is noteworthy is that the dictionary user may still get information regarding the morphology of lemmata as word stems are frequently indicated through bold font while prefixes remain in regular font. Thus for the word *abafundisi* (learners), class 2 prefix *aba-* is in regular font while the stem *-fundisi* is in bold. Unfortunately, this is not done consistently. For example, immediately under the lemma 'abakhi' (builders) in which regular and bold fonts respectively distinguish between the prefix and stem, no similar distinction is made regarding the lemma **abakho** (yours) which is entirely in bold while *aba-* in the same word is a prefix as it is the case in *abafundi* (learners) and *abakhi* (builders). Although it appears that the distinction between prefixes and stems is made regarding nouns, lexical items belonging in the speech-type categories such as possessives, adjectives, relatives, etc. also display the prefix-plus-stem structure, thereby making the distinction made regarding nouns also relevant.

A possible criticism of the word lemmatisation method adopted for the *Oxford Bilingual School Dictionary: IsiXhosa and English* is that it results in the inclusion of closely related words which leads to unnecessary redundancy, assuming that once one knows the meaning of *-funda* then the meaning related words such as *umfundi* (learner), *abafundi* (learners), *fundisa* (teach), *umfundisi* (teacher/preacher), etc. would be predictable. Such criticism is amplified by the economics of dictionary-making which includes limited but precious time and space. However, it is important to quell such concerns by recalling that this dictionary is based on a corpus and is aimed at specific users (learners) whose use will be linked to specific situations (school-work). The corpus-based approach to the various lexicographic tasks involved in the compilation of this dictionary and the fact that this approach did not exclude items which are less frequent or even absent in the corpus indicate that the compilation of the dictionary never lost sight of balance between the economisation of the precious lexicographic resources and the satisfaction of user needs. That the user who may not establish the lexical relations intuitively may get information that may be considered redundant should in this case be considered as sensitivity towards the inexperienced language learners/users and dictionary users. Such sensitivity becomes a necessary condition for user-friendliness in relation to dictionaries.

Unlike the isiXhosa wordlist, the English wordlist of the *Oxford Bilingual School Dictionary: IsiXhosa and English* does not present novelties that require introduction to anybody who has previously used an English dictionary. Any user who has mastered the alphabet, as it may be expected, even though not

guaranteed, of the target users of this dictionary, should be able to locate all the included words with relative ease. Attention should perhaps be paid to the microstructural aspects of the English–isiXhosa side which differ from those of the reverse side. In addition to the type of speech information, isiXhosa translation equivalents, grammatical data that applies to respective lemmata, e.g. singular/plural forms for nouns or tenses for verbs, as well as cross-references where applicable and pronunciation data form part of the obligatory microstructure for English lemmata. The pronunciation information is not provided by means of the International Phonetic Alphabet (IPA) transcription as is the case in most dictionaries but in a manner that is intended to help the learners who may not be familiar with the IPA. For example, with respect to the word *acquire*, the user is told to "say uh-**kw**y-uh" (p. 239). While this could be seen as a worthwhile departure from the use of the IPA symbols, whether the non-mother-tongue learners of English, who have isiXhosa as their mother-tongue in this case, will not pronounce *uh-* in the same way as they pronounce *u* in their mother-tongue is an issue that could be determined through dictionary user-research. At this stage, the lexicographers may be commended for their effort which goes beyond adopting this approach to pronunciation indication but also for the guidance that they provide regarding this approach (p. xiii).

Besides the microstructural data that is provided in relation to different types of lemmata in both the isiXhosa and English wordlists, constituting single articles, the dictionary provides additional information in the form of complex articles (Gouws and Prinsloo 2005: 91-92). The most obvious of these in the *Oxford Bilingual School Dictionary: IsiXhosa and English* consist of texts that are boxed and highlighted at the end of the simple articles. For example, after giving the relevant data regarding the second lemma sign **ache**, the user is advised to "[n]ote that *iingqaqambo* and *iintlungu* are used in the plural sense. Sentences that use the noun *ache* are often translated using the relative *-buhlungu* in isiXhosa" (p. 239). In a similar way, a note is made under the lemma 'ilantshi' that "[a] more formal term for *lunch* is *isidlo sasemini*" (p. 69). That way, the lexicographers are able to extend the supply of data beyond the default article structure constituting the obligatory microstructure, thereby drawing the attention of the user to critical nuances regarding the usage of a particular lexical item.

The final aspect of the two wordlists that is highlighted is the use of illustrations in the *Oxford Bilingual School Dictionary: IsiXhosa and English*. Illustrations seem to be used to elucidate the paraphrases of meaning, but some of them also appear to increase the cognitive value of the dictionary. For example, the illustration of the plant under the lemma 'isityalo' will not only give a learner a picture of how a plant looks like but also the labelled different parts of plants (p. 112) which constitute important knowledge in the natural sciences curriculum. The same goes for the illustration provided to supplement the description of **pivot** in relation to science and technology (p. 446).

Besides the main text of the dictionary, which comprises the two wordlists

discussed above, the user will learn a lot from the outer texts of *Oxford Bilingual School Dictionary: IsiXhosa and English*. In the front matter, texts such as 'Impawu zesichazi-magama/Dictionary features', 'Intshayelelo/Introduction' and English pronunciation provide useful background information regarding the dictionary or its aspects. Such information has the potential to provide the much needed dictionary pedagogy, especially among speakers of African languages, such as the isiXhosa-speaking target users of this dictionary. Users are oriented on the use of this dictionary, the experience of which may be useful in their search for knowledge in other lexicographic products. The same applies to 'Imisebenzi yesichazi-magama/Dictionary activities' and 'Indlela yokusebenzisa isichazi-magama/How to use your dictionary' in the middle matter. Texts such as 'Incwadi youshishino/Formal letter', 'I-imeyile youshishino/Formal email', 'Upelo lweSMS/Text messages or SMSes', among others, will enable the dictionary users to produce texts that are appropriate in specific situations, while the various texts dealing with some aspects of isiXhosa and English will increase the user's knowledge regarding the respective languages. That way, the lexicographers go far enough in trying to ensure that, as its genuine purpose, the dictionary helps "people learning isiXhosa to understand and use isiXhosa better" and those who are "learning English to understand and learn English better" (p. vi).

In light of the above, the publication of the *Oxford Bilingual School Dictionary: IsiXhosa and English* can be celebrated as a great milestone in the lexicography of isiXhosa and can be expected to have a positive impact in the learning of both isiXhosa and English. Obviously, this review is too brief to proclaim itself as comprehensive enough to provide an exhaustive commentary of this new dictionary. However, it hopes to have given major insights regarding this dictionary, but more comprehensive and in-depth studies remain relevant in order to expand knowledge regarding isiXhosa lexicography and key issues in relation to school dictionaries in general. For now, we can celebrate its publication and encourage the target users to use it to its maximum potential. Even for university students of isiXhosa, it may be considered as an upgrade rather than a downgrade, given that the popularly used Fischer's *Oxford English Dictionary* has been printed for not less than twenty-one times but without a single revision.

References

- De Schryver, G.-M. 2010. Revolutionizing Bantu Lexicography — A Zulu Case Study. *Lexikos* 20: 161-201.
- De Schryver, G.-M. 2008. A New Way to Lemmatize Adjectives in a User-friendly Zulu-English Dictionary. *Lexikos* 18: 63-91.
- De Schryver, G.-M. et al. (Eds.). 2014. *Oxford Bilingual School Dictionary: IsiXhosa and English*. Cape Town: Oxford University Press Southern Africa.
- Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PRess.

Gouws, R.H. and D.J. Prinsloo. 2005a. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. *International Journal of Lexicography* 18(1): 25-46.

Kropf, A. 1899. *A Kafir-English Dictionary*. Alice: Lovedale Press.

Dion Nkomo
School of Languages
African Language Studies Section
Rhodes University
Grahamstown
South Africa
(d.nkomo@ru.ac.za)

Mariusz Kamiński. *A History of the Chambers Dictionary*. Lexicographica Series Maior 143. 2013, xii + 297 pp. ISBN 978-3-11-031250-8. e-ISBN 978-3-11-031273-7. ISSN 0175-9264. Berlin/Boston: Walter de Gruyter. Price: €84.

As there have been few attempts at a systematic study of the history of popular general dictionaries, this book on the *Chambers Dictionary*, whose tradition goes back to the nineteenth century is especially worthy of cognisance. This study by Kamiński traces the development of the *Chambers Dictionary* through its successive editions: the three editions which appeared during the last thirty years of the nineteenth century, the first dating from 1867, the nine editions which were published during the twentieth century, the first in 1901, and the second about fifty years later in 1952, and the four editions originating from the first eleven years of the twenty-first century. This comparative approach gives scholars and students insights into the procedures followed by the *Chambers* compilers, the aims they wished to achieve and the problems they encountered during the consecutive revisions of the dictionary.

After explaining his way of numbering the editions and his method of sampling to ensure a consistent comparison of the editions, Kamiński expounds in Chapter 1 the methodological and theoretical foundations of his analysis of the structure of the successive editions of the dictionaries: the arrangement of entries, the selection of morphological forms, the choice and growth of vocabulary, defining meaning, the order and discrimination of senses, etymology, syntagmatic and paradigmatic information, pronunciation and outside matter. He discusses each of these points in the context of the original edition in Chapter 5, and of the subsequent editions in Chapters 6–15. The framework of the analysis has been based partly on Hausmann and Wiegand (1989) and their followers, for example Frączek (1999), in which one of the recurring themes is the major principle of the theory of lexicographical functions, according to which dictionaries are considered as objects of use meant for a specific group of users (Tarp 2008; Wiegand 1987).

Chapter 2 presents an outline of the biographies of the Scottish brothers William and Robert Chambers, the founders of the W & R Chambers publishing house. Robert died in 1871, a year before the publication of the second edition of the *Chambers Dictionary*, and William in 1883, after which Robert Chambers (1832–1888), William's nephew, became the head of the firm. Since the 1993 edition, the *Chambers Dictionary* has been published by Chambers Harrap.

By describing the socio-historical factors which stimulated the growth of educational publishing in Victorian Britain, Kamiński deals with the origin of the dictionary in Chapter 3, also discussing the sources of the dictionary by referring to the intellectual ideas, which might have influenced the structure of the original edition. These sources represent the overlapping of two traditions, the earlier one of philosophical etymologising, and the new one of the historical-comparative philology.

Chapter 4 provides a biographical account of the editors involved in the compilation of the different editions of the *Chambers Dictionary*: James Donald was responsible for the first two editions of 1867 and 1872, Thomas Davidson was the editor of the next two editions, the last edition of the nineteenth century in 1898 and the first edition of the twentieth century in 1901, William Geddie edited the 1952 and 1959 editions, Agnes Macdonald, also two editions, those of 1972 and 1977, while Betty Kirkpatrick edited the edition published in 1983. Catherine Schwarz was managing editor for the editions of 1988 and 1993 and, together with Robert Allen, was a consultant editor for the 1998 edition. For the editions of 2003 and 2006, Ian Brookes was the editor and for those of 2008 and 2011, Vivian Marr. Kamiński observes that recent *Chambers* lexicographers had a more professional linguistic background, being better qualified for compiling dictionaries than the early editors, whose academic training was unrelated to lexicography.

In Chapter 5, a detailed analysis of the design of the first edition of *Chambers* is presented within the framework expounded in Chapter 1. This edition is given special attention because it serves as a point of departure for the discussion of subsequent editions. Designed chiefly as a school dictionary for educational purposes, this edition of *Chambers*, in spite of its title, devoted relatively little space to etymological information, but etymology nevertheless played an important role in relation to meaning, definitions and entry structure which is examined in the first part of the chapter before the other descriptive and pedagogical considerations are dealt with.

The evolution of the dictionary is the subject of Chapters 6–15, each analysing how a particular aspect of the dictionary listed in Chapter 1 has changed over the years. The thematic approach followed in these chapters ensures a proper perspective on the changes in the consecutive editions. Each chapter is followed by a summary of the main points. Tables schematically presenting the different thematic aspects are usually given at the end of the relevant chapters. The conclusion of Chapter 6, dealing with the optimal arrangement of entries shows that editors experimented with various presentations, beginning with an etymological grouping, then a strictly alphabetical arrangement and finally nesting, which was used throughout until the latest editions.

Concerning the selection of morphological forms, Chapter 7 shows that it was not until the third edition that the dictionary could cover more vocabulary thanks to extensive nesting where main entry status was given to self-explaining derivatives and compounds in the earlier editions. Whereas certain morphological forms, such as combinations, and prefixes and suffixes, were treated in separate lists in the back matter of earlier editions, these were slowly moved to the central list in later editions.

Chapter 8 reveals the gradual growth in the dictionary's range of vocabulary, reflecting the changing image of the intended user: from the educational aim of its first edition to the broader scope of a general audience. The number of characters in the main body of the dictionary increased five times, from three

million in the first edition to nearly fifteen million in the 2003 and later editions.

Chapter 9 discusses the development of the definitions with a view to style, use of keywords and defining vocabulary. Focusing on major definition types, it investigates how these have changed quantitatively since the first editions. Separate sections deal with encyclopaedic and idiosyncratic definitions. Throughout the history of the dictionary, definitions have remained traditional.

The order and discrimination of senses are considered in Chapter 10. In all editions, a strong preference for a historical ordering has been shown. Individual senses are separated by using either a colon or a semicolon, the editors, except in the first edition, not trying to develop this system further by distinguishing larger categories of senses or dividing them into sub-senses. The use of these traditional indicators implies the view of the existence of a clear-cut semantic structure, with rigid boundaries between and no overlapping of meanings.

In Chapter 11 modifications in etymology are discussed. Significant changes in the third and fifth editions were brought about by Davidson and Geddie respectively. Except for the correction of etymologies, hypothetical information that was unsupported by the comparative-historical method was deleted. Since users often found it difficult to interpret etymologies in the early editions, revisions in subsequent editions were aimed at clarifying the interpretation of punctuation marks used in etymologies. Etymologies the first editor added to a number of complex and compound forms were restricted by subsequent editors to the entries for bases only.

Chapter 12 shows that the system of usage labels has expanded both quantitatively and qualitatively. Each successive edition employed a more developed system of marking. While more labels were used, their proportions also changed: whereas there were relatively fewer source labels, labels indicating time/frequency, attitude, register/style and field increased significantly. This testifies not only to the rapid development of science and the emergence of new areas of human interest, but also to the editors' increased lexicographical competence.

As shown in Chapter 13, the *Chambers* editions provided little syntagmatic and paradigmatic information. Syntagmatic information is restricted to the indication of word-class and the sub-category (transitive or intransitive) of a verb as well as the selection of set expressions, such as prepositions complementing verbs, direct objects of verbs, syntactic phrases and citations. Paradigmatic information appeared chiefly in the form of antonyms, which was not provided on a regular basis.

Chapter 14 deals with the pronunciation system used in *Chambers*, based on respelling with diacritics over vowels, and different typeface for distinguishing between certain consonants. Revisions effected in later editions aimed at making it more phonemic, developed and precise. In all editions, the editors used respelling systematically for all headwords, and, in addition, also for some subheads. With the expansion of the dictionary, the treatment of pronun-

ciation was rethought, especially with regard to phonetically predictable terms.

In Chapter 15 the outside matter of the *Chambers* dictionaries is discussed, and the contents of the front, middle and back matter considered. The user's guides in the front matter of early editions were less detailed than those in later editions. In the earlier ones, the guides were structured very similarly, containing several usually short sections explaining the arrangement of entries and senses, and containing information on pronunciation and etymology. Except in some early editions, the middle matter in most editions was little developed, although from the 1983 edition onwards, the middle matter contains occasional lists of derivatives given at the foot of relevant pages. In all editions, the back matter incorporates a number of lists displaying encyclopaedic and linguistic types of information, the earlier editions mostly containing linguistic information, and the later ones showing an increasing amount of encyclopaedic information.

These chapters are followed by a few pages containing concluding remarks. Kamiński comments that his impression is that the greatest changes were made in the editions published in the nineteenth century (1872, 1898) and the early years of the decades of the following centuries (1952, 1972, 1983, 1993, 2003). He distinguishes two major lines of development, one regarding the comprehensiveness of coverage, and the other regarding user-friendliness.

The reorganisation of entries by means of extensive nesting in the third edition helped editors to use space more efficiently, making room for new words. Space was also saved by deleting redundant types of information and presentation, such as regularly inflected forms (since the 1898 edition), pictorial illustrations (since the 1952 edition) and punctuation in the etymologies (since the 1993 edition). Because of the space restrictions, *Chambers* also did not systematically include illustrative examples.

Although attempts have been made to produce a more user-friendly dictionary, for example by ordering senses according to the usage principle, employing more transparent labels, using punctuation marks more consistently in the etymology, dividing the entries into sections, giving more information on the use of the dictionary, and simplifying the difficult language of the definitions, certain features remained traditional, for instance brief and precise definitions sometimes at the expense of clarity, defining by reproducing morpho-semantic elements, clustering lexical items under the base word, using alternatives, enumerations and near-synonyms as definitions, the inability to delineate clear meaning boundaries and the prototypically organised structure of lexical categories.

However, in the print version of a dictionary, it is difficult to achieve the goals of comprehensive coverage and user-friendliness simultaneously. These restrictions have been overcome in many ways in the electronic version of *Chambers*, for instance, by presenting single senses on separate lines, by using full forms of labels, and by providing pronunciation audibly in combination with respelling.

In the Appendices appear examples of pages selected from the early editions published in 1867, 1872, 1898 and 1901. These are followed by Samples 1 and 2 containing excerpts from all editions of *Chambers*. Sample 1 represents the macrostructure, comprising six sections of the word-list, while Sample 2 reflects the microstructure, covering a selection of full entries. The samples were analysed in detail for statistical purposes. Throughout the analysis cross-references pointing to particular entries in Sample 2 are made. When necessary, Kamiński sometimes also considered entries from other parts of the dictionaries, especially when a particular feature did not occur in the samples.

The bibliography is divided into three sections: dictionaries (editions of the *Chambers Dictionary*, in chronological order, and other dictionaries and encyclopaedias mentioned in the text), Internet sources, and other literature referred to in the discussion. The book is concluded by an index.

From the preceding discussion, it has become clear that the book is a systematic, thorough and exemplary study of the *Chambers Dictionaries*. Scholars and students interested in the history of these dictionaries will discover much valuable and useful information in the analyses, and those intent on undertaking a similar study of the development of the same kind of dictionary will find it an excellent model for approaching such research.

References

- Frączek, A.** 1999. *Zur Geschichte der deutsch–polnischen und polnisch–deutschen Lexikographie (1772–1868)*. Tübingen: Max Niemeyer.
- Hausmann, F.J. and H.E. Wiegand.** 1989. Compound Parts and Structures of General Monolingual Dictionaries: A Survey. Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (Eds.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie / Dictionaries. An International Encyclopedia of Lexicography / Dictionnaires. Encyclopédie internationale de lexicographie*: 328–359. Berlin: Walter de Gruyter.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 1987. Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. *Lexicographica. International Annual for Lexicography / Internationales Jahrbuch für Lexikographie / Revue internationale de lexicographie* 3: 178–227.

J.C.M.D. du Plessis
Bureau of the Woordeboek van die Afrikaanse Taal
Stellenbosch
South Africa
(jcmddp@sun.ac.za)

Taasisi ya Taaluma za Kiswahili (TATAKI). *Kamusi ya Kiswahili Sanifu*. Third Edition. 2013, xvi + 687 pp. ISBN 978-019-573820-9. Nairobi: Oxford University Press. Price: Kshs. 840.00 (Kenyan Shilling).

A monolingual dictionary is compiled to cater for the speakers of a particular language. The need for this kind of dictionary emanates from the wider use of the language concerned and introduction of new words into that language which may give problem to the speakers. The importance of a dictionary also increases when the language is rendered into writing with different texts expressing various fields of knowledge through it. This scenario may confuse several speakers of that language who may have inadequate vocabulary to enable them understand the meaning of words in different texts. This is the scenario which encountered the Kiswahili speakers when the *Kiswahili-Kiswahili dictionary* was compiled by Fredrick Johnson (1935). By 1970 it was realized that Johnson's dictionary had some inadequacies because it could not cope with the faster development of Kiswahili. A new dictionary had to be compiled and published in 1981 by The Institute of Swahili Research — *Taasisi ya Uchunguzi wa Kiswahili (TUKI)* to cater for this paucity.

A decade later, the *TUKI* (1981) also faced challenges of embracing the new developments in Kiswahili, hence a need for another dictionary to cater for these challenges arose. *Kamusi ya Kiswahili Sanifu* (2013) henceforth KKS third edition is a product of up-to-date linguistic research done by several Kiswahili scholars from diversified fields in Kiswahili scholarship. These scholars as recommended by Cowie (1983: 135) have carefully synchronized pedagogical consideration with linguistic factors in the design of KKS. This is a general dictionary in which the lexicographers have made information maximally explicit in the text and especially through exemplification. The front matter of KKS has a list of lexicographers who have participated in different panels in the development of KKS since 1964–2012; preface written by Dr. Aldin Mutembei which underscores the need for this current edition. The introduction section informs the user about new entries such as etymology of words, some verb extensions, sense relations, noun classification through morphological, syntactic and semantic approaches, pronunciation and International Phonetic Association (IPA) alphabet which were not covered in earlier editions. KKS contains more than 285,000 headwords, of which 25,000 are synonyms and verb extensions, 2000 new headwords not existing in the previous editions and more than 1150 attractive pictures and drawings.

The section on the guide to the user (x-xvi) has a summary of Kiswahili grammar and an illustration on how to use KKS. The summary of Kiswahili grammar in KKS shows grammatical elements of Kiswahili which may confuse a user if not explained properly. Such elements as illustrated through KKS are:

- i. Components of word such affixes, compound words;
- ii. Word categories such as noun, adjectives, verbs, adverbs, demonstratives etc;

- iii. Grammatical agreements;
- iv. Grammatical functions such as noun classes, types of verbs: transitive and intransitive, tense, mood;
- v. Verb extensions;
- vi. Punctuation marks.

Headwords have been entered from page 1 to 656 of KKS. The entry has been done alphabetically. Every entry of a headword has been followed by grammatical information which includes phonological and syntactic information. The information on scope and stylistics with reference to a particular headword has also been given. After information on grammar, KKS gives an explanation on meaning of every headword followed by examples on how that headword can be used in the composition of various sentences in different contexts. KKS has also shown a standard orthography/spelling wherever there is one headword with different spelling. For example:

ilimradi/ilimradi/ pia **alimradi**/alimradi/ ku kwa masharti kwamba;
iwapo, mradi.
angalau/angalau/ pia **angaa**/anga:/, **angalao**/angalao/ ku neno litumikalokueleza jambo hilo bora zaidi; **walau, falau.**

These examples show that the orthography preceded by the word *pia* is not a standard one but is used in different context.

KKS depicts several grammatical information such as the word category in which the headword follows into, noun class, case, verb extension etc. KKS has used abbreviations to capture this information in Kiswahili: kt (kitenzi-verb), kl (kielezi-adverb), kv. (kivumishi-adjective), kw (kiwakilishi-pronoun), ku (kiunganshi-conjunction), nm (nomino-noun), ele (elekezi-transitive), sie (sielekezi-intransitive) etc. This information is found in the front matter of KKS.

The verb structure has been distinguished in KKS by separating the radical from affixes by a dot as illustrated by the following example from KKS: *pembu.a, pen.a, pend.a, shak.a, shajilish.a*, etc. The problem with this format of verb extension is that naïve speakers of Kiswahili may follow it to for some verb extensions which are not desirable in Kiswahili as observed by Mdee (2010: 80).

One aspect of language elaboration which is quite interesting in KKS is the inclusion of very recent vocabulary in Kiswahili as its headwords. These vocabularies reflect recent trends in scientific, economic and social development in human society. Such vocabularies are: *tuktuk, tishu, intraneti*, etc. There is also a section dealing with synonyms which is from page 647 to 656. This section is also arranged alphabetically. KKS has come up with colored pictorial presentation of internal parts of human body, attires, and types of fish and other marine creatures, animals and insects, varieties of trees and crops,

horticultural produce, musical instruments, war weapons, tools of different trades such as carpentry, masonry, parts of motorized engines all presented from page 657 to 672. Furthermore, there is some note on grammar again from page 673 to 675. KKS has done a good job by presenting a list of names of countries and nationality in English and a translation of the same in Kiswahili from page 676 to 680. Finally KKS has given a very detailed presentation on grammar from page 681 to 687.

In my considered opinion this is a very good dictionary for any person who would like to learn Kiswahili and also Kiswahili experts.

References

- Cowie, A.P. 1983. *English Dictionaries for the Foreign Learner*. Hartmann, R.R.K. 1983. *Lexicography: Principles and Practice*: 135-152. London/New York: Academic Press.
- Johnson, F. 1935. *Kamusi ya Kiswahili, yaani Kitabu cha Maneno ya Kiswahili (Swahili Dictionary, i.e. a Swahili Wordbook)*. London: Sheldon Press.
- Mdee, J.S. 2010. *Nadharia na Historia ya Leksikografia*. Dar es Salaam: Oxford University Press.
- TUKI (Taasisi ya Uchunguzi wa Kiswahili). 1981. *Kamusi ya Kiswahili Sanifu*. Dar es Salaam: Oxford University Press and Institute of Kiswahili Research.

Stephen Oluoch
Department of Languages
Kisii University-Kenya-East Africa
Kenya
(yahuma1973@yahoo.com)

Sylwia Wojciechowska. *Conceptual Metonymy and Lexicographic Representation*. 2012, 209 pp. ISBN 978-3-631-63510-0. Frankfurt am Main: Peter Lang. Price: €53.15.

The field of lexicography has proven a fruitful ground from which language learners can benefit (Boers and Lindstromberg 2006). The cognitive approach to dictionary making relies on the motivation of a word's senses so that the connection between literal and nonliteral meaning can be elucidated by foreign language students more easily. Geeraerts (1990), one of the main proponents of cognitive lexicography, claims that many dictionaries make use of cognitive principles such as the prototypicality of senses and their network-structured nature. Experimental studies on the acquisition and retrieval of vocabulary, for instance, have shown the benefits of this approach (Boers 2000). Unlike studies on metaphor encoding (e.g. Van der Meer 1997, 1999; DeCesaris and Alsina 2002; Philip 2006), Wojciechowska's (2012) book and dissertation are among the first attempts to uncover the (un)systematicity of metonymic shifts in English dictionaries. The author's aim is 'to show how the tenets of the cognitive theory of metonymy can benefit the representation of metonymic lexemes in pedagogical lexicography' (2012: 16).

The book is composed of four chapters which encompass a theoretical overview of metonymy and an analysis of the microstructure, including definitions, examples and codes, of five English Monolingual Learner's Dictionaries (MLD hereafter). They are, as cited in Wojciechowska (2012: 167), the so-called 'Big Five', namely the *Cambridge Advanced Learner's Dictionary* (2005), the *Collins COBUILD English Dictionary for Advanced Learners* (2003), the *Longman Dictionary of Contemporary English* (2003), the *Macmillan English Dictionary for Advanced Learners* (2007) and the *Oxford Advanced Learner's Dictionary of Current English* (2005). Her present work was previously published in a shortened version in *Proceedings of the 14th EURALEX International Congress* as Wojciechowska (2010). Chapter 1 provides a theoretical account of metonymy, highlighting the most prominent notions within the field, including the idea of domain highlighting, conceptual contiguity, the semantic elaboration of the source and the interplay of metonymy with metaphor. Chapter 2 deals with the organization of metonymic senses in the microstructure of the dictionary, including metonymic chains, and proposes an ordering according to our conceptualization of meaning which goes from literal to figurative senses. Chapter 3 analyzes the *definiens* of metonymic senses according to the effectiveness and coherence of the strategies followed in the 'Big Five'. Finally, chapter 4 looks at the lexicographic coding of count-mass nominal metonymies and points out the lack of transparency of the grammatical markers for distinguishing them.

'Metonymy in Cognitive Linguistics' (43 pages) presents an overview of the Cognitive Theory of Metaphor and Metonymy (Lakoff and Johnson 1980; Lakoff 1987; Lakoff and Turner 1989) together with foundational definitions of metonymy and domain. From the canonical definitions in Lakoff and Johnson

(1980) of both metaphor and metonymy, we are walked to those offered respectively by Barcelona (2000) and Croft (2002) in subsection 1.2: two-domain and one-domain correspondences in the former are equated with superordinate-to-superordinate and subordinate-to-superordinate correspondences in the latter. Wojciechowska reports as well Nerlich and Clarke's (2001) pragmatic characterization of metaphor as extracting new information from old words and of metonymy as extracting more information from fewer words. Finally, Radden and Kövecses' (1999) Idealized Cognitive Model-based mental access provided by the source to the target is cited as an essential add-up to the cognitive understanding of metonymy. Subsections 1.3 and 1.4 establish a comparison between metonymy and ellipsis firstly, and metonymy and synecdoche secondly. In short, ellipsis works at the linguistic level whereas metonymy works at the conceptual level. The author follows Seto (1999) in his distinction between synecdoche and metonymy, opposing the conventional classification of the PART FOR WHOLE synecdochic pattern as metonymic in the literature (cf. Radden and Kövecses 1999).

A series of problematic issues are presented in subsection 1.5, entitled 'Problems with the Standard Definitions of Metaphor and Metonymy'. Domains are still not defined with precision, at least in the current cognitive linguistic theory. This fact represents a hindrance to the identification of what constitutes metonymy or not as well as to the identification of the nature of mappings between entities. Several proposals elaborating on the Langackerian notion of domain have been propounded, among which the conceptions of matrix domains (Croft 2002) and superordinate domains (Ruiz de Mendoza Ibáñez 2000: 116) stand out. The former option, that of domain matrices, is used by Wojciechowska in her dissertation. In 1.6 alternate solutions are brought forth, namely domain highlighting (instead of domain mapping) and domain reduction and expansion. In this respect, the author favours the idea of semantic elaboration to explain metonymy. Among those alternates, she downplays in 1.7 the role of conceptual contiguity and the Langackerian notions of facetizations and zone activation (Langacker 1987). Domains and conceptual contiguity, following Peirsman and Geeraerts' (2006) critical view, is substituted by the characterization of metonymies according to their prototypicality, strength of contact and boundedness. The distinction made by Paradis (2004) between metonymy, on the one hand, and facetization and zone activation, on the other hand, is disregarded by Wojciechowska because if such were the case, no metonyms would be listed in dictionaries.

In consequence, semantic elaboration overruns other options because, in her own words, it 'is not based on the vague notion of a domain [...] does not incorporate mapping' and also includes 'the backgrounded source meaning in the conceptual structure of the target meaning' (2012: 40). Additionally, she adopts the formula X-Y instead of the pervasive X FOR Y to indicate the absence of substitution of the source by the target (2012: 18). In *I like Mozart*, the target *Mozart* has not substituted simply 'music' but 'music by Mozart' (Warren

1999: 127-128) and similarly, in *She is just a pretty face, a pretty face* does not substitute the whole person as beautiful but only her beautiful face belonging, naturally, to her whole body (Radden and Kövecses 1999: 19). Next, subsection 1.8 comprises an analysis of different typologies of metonymies as proposed by Radden and Kövecses (1999), Seto (1999), Peirsman and Geeraerts (2006) and Ruiz de Mendoza Ibáñez (2000). A point of agreement among them all is the basic nature of the WHOLE-PART pattern from which others stem. The final remark in chapter 1 comprises the interaction of metaphor and metonymy in many real-language instances. As stated in subsection 1.9, it may be more realistic to talk of a continuum with each of the two conceptual operations at opposite ends and a series of 'in-between' cases rather than separate notions with strictly defined boundaries.

Chapter 2 (35 pages) initially overviews the theoretical stances towards the disambiguation of words' senses conceived as the first lexicographic step. Once dictionary makers know what senses to include, it is time to worry about what ordering they should adopt within the entry. Wojciechowska's first empirical study and its subsequent results attempt at checking whether conventionalized metonymic senses are included within the 'Big Five' and how they are included, i.e. if dependently or independently from other senses. Her stance in this chapter is that 'as institutionalised metaphors are covered in MLDs, conventionalised metonyms also deserve their share of dictionary space' (2012: 64).

Subsection 2.1, 'Word Sense Disambiguation', summarizes two distinct stances regarding the notion of 'sense', a radical one where dictionaries only contain potential meaning and a milder one that follows the dictates of Rosch's theory of prototypes (Ungerer and Schmid 1996). The first trend gives a preponderant role to context. In this respect, Cruse's (1986: 50-54) distinction between 'sense selection' and 'sense modulation' is exemplified by the words *bank* and *cousin* respectively. Among *bank* as 'financial institution' and *bank* as 'side of a river', which are ambiguous senses, we select the sense that is appropriate to a specific context. Conversely, in the vague word *cousin* we modulate or further specify the sense of cousin as female or male without changing its basic characterization. One of the main ideas that Wojciechowska propels is that of a gradual nature in meaning understanding, be it metaphor and metonymy, ambiguity, polysemy or vagueness. Therefore, context bears, according to this view of meaning as potentiality, a crucial role in understanding the sense of a word (Kilgarriff 1997; Hanks 2000). In lexicographic production, corpora are seen as the most reliable sources for obtaining those context-dependent examples. Nonetheless, the author does not make use of any corpora but treats dictionaries as corpora-like sources for language in her study.

The second stance aims at listing in dictionaries the most prototypical senses of each lexeme, thus still considering, though to a lesser extent, a word's context in discourse. The processes of 'lumping' and 'splitting' denote what in lexicography correspond to these two diverging stances on the inclusion of

word sense in lexicographic works: the former, that of potentiality of meaning, minimizes the number of senses and the latter, that of prototypicality, maximizes them (Adamska-Sałaciak 2006; Atkins and Rundell 2008). Wojciechowska banks on the idea of coherence with respect to the strategy used to list words' senses in dictionaries. Furthermore, she propounds 'template entries' as 'a framework designated to facilitate writing entries for words that belong to lexical sets' (Atkins and Rundell 2008: 123 as cited by Wojciechowska 2012: 68). Her statement that 'it would be an exaggeration to say that there are no word meanings outside context' (2012: 68) and her belief that only conventionalized metonymies should be listed in dictionaries are an 'in-between' stand with respect to the approaches elaborated above.

Subsection 2.2, 'Sense Ordering', makes reference to the three existing types of ordering, namely the historical, logical and frequency-based ones. Precluding the first one to historical dictionaries like the OED, the logical ordering presents the 'core' and prototypical senses first following the second approach to dictionary making developed in subsection 2.1. The ordering according to frequency, in turn, is based on corpora studies and presents the most used words first. This subsection 2.2 tackles next the cognitive ordering of senses which, aware of the linearity problem (Geeraerts 1990) and unable to express the complex networks of meaning representing words in a two-dimension volume, embraces a series of tenets. Following Van der Meer and Sansome (2001), the literal meaning must precede the figurative meaning. Similarly, Wojciechowska believes, as Mel'čuk (1988) does, that metonymic senses must precede metaphorical ones. Finally, the source of a metonymy should come before the target (2012: 74). Another issue raised in 2.2 is 'metonymic chains' and 'double metonymies' which the author names 'multiple metonymies' to avoid the implication that all of the targets of a metonymic source must be related or further elaborated. For example, the different targets of the metonymic source *silver*, namely 'coin', 'utensil/object', 'colour' and 'medal', are not semantically elaborated.

The empirical study in 2.3 that the author presents is motivated by the scant number of studies analyzing metonymies in dictionaries vis-à-vis the more numerous studies of the inclusion of metaphors in dictionaries. A hundred metonymic senses grouped under the metonymic patterns PART-WHOLE, WHOLE-PART and PART-PART proposed in Radden and Kövecses (1999) are studied. They all are conventionalized, nominal, frequent (at least present in two of the dictionaries) and not outdated metonymies. Additionally, they are cases of regular polysemy because their pattern of meaning extension is repeated in more than one linguistic unit (Apresjan 1995). The results yield a very heterogeneous panorama in the treatment of metonymy within the 'Big Five'. CALD2 is the only dictionary that presents metonymic sources and targets as distinct entries due to its policy of homonymy maximization. It is also unique in its use of brackets for metonymic senses. OALDCE7 and LDOCE4 list metonyms as separate senses of a lexical item. In contrast, MEDAL2 opts for

listing the metonymic target as a subsense in accordance to Wojciechowska's policy. Broadly speaking, 'LDOCE4, MEDAL2 and OALDCE7 stand out as more successful than either CALD2 or COBUILD4' (2012: 85) in terms of how metonymic senses are treated in the 'Big Five'. Regarding the ordering of the source before the target, however, over 80% of metonymies are successfully represented from the cognitive point of view.

In subsection 2.4 a discussion of the results is proffered, some facts of which are highlighted here. Firstly, there is a lack of cohesive treatment of metonymies within dictionaries and in each dictionary individually. For example, *lamb* in the pattern ANIMAL–MEAT (WHOLE–PART) is listed in the dictionaries using strategies such as the conjunction *or*, two senses with the source first or the target as a subsense of the source. The most coherent dictionary with respect to the representation of specific metonymic patterns is OALDCE7. Secondly, the source and both source and target are sometimes, though rarely, omitted. An example of the source omission is *stiletto* in its 'heel' meaning in CALD2 and COBUILD4. The full omission of metonymic source and target happens with the VOICE–PART pattern in *alto*, *bass* and *tenor* in COBUILD4. Lastly, the ordering of source and target is mostly followed in all dictionaries, except for some internal variation in specific patterns. Some cases become even more complicated when a lexical item such as *hand* has both metaphorical and metonymic senses: in many cases its metaphorical extension is placed between the metonymic source 'hand' and the target 'manual worker' illustrated by the BODY PART–PERSON pattern.

Chapter 3 (37 pages) assesses the 'Big Five' in relation to their definitions of metonymic senses. The difficulty to distinguish and clearly define them intersects with the central role of definitions, together with examples, as tools facilitating the users' understanding of meaning. The cognitive approach to lexicography overshadows the structuralist tradition as evinced in chapters 1, 2 and 3 of Wojciechowska's book. From this perspective, encyclopaedic information is prone to be reflected in the entries when necessary to make the meaning and its connections clearer. Comprehensibility and clarity are hence seen as decisive factors in the quality and success of the definitions in MLDs.

Among the defining strategies presently in the field, the canonical analytical definition and the full-sentence definition are pervasive in the 'Big Five'. Their presence is solidly grounded since they are able to accommodate both source and target in the definition as opposed to other strategies which are shorter. Wojciechowska's results confirm these facts by which 'the examination of the defining styles for the fifty metonyms in the present enquiry has revealed that the analytical definition predominates' (2012: 106). Within those, the target appears first in the function of *genus proximum* and the source appears as the explanation or *differentia specifica*.

Six issues with defining are found by Wojciechowska in her analysis, of which the second bears no repercussion to the successfulness of the definitions. Firstly, there is an inconsistent use of reference strategies illustrated by the

definition of *condominium*. CALD2, COBUILD4 and MEDAL2 use the definiendum, LDOCE4 uses the controlled defining vocabulary and OALDCE7 uses anaphoric reference. The author believes that 'template entries' would yield a homogeneous definition of metonymic senses in each dictionary. Unexpectedly, the use of derivative and base words is not condemned by the author, as Landau (2001) does, if those words are defined somewhere else in the dictionary. For example, the verb–noun conversion of *dance* showing the metonymic pattern ACTION–EVENT seems acceptable to the author: an alternative definition in the fashion of 'a social event where people move their bodies to music' (2012: 121) would be longer and more difficult to understand for beginners. Thirdly, the semantic relation between source and target is based on how close the defining vocabulary of each of them is to the other. Wojciechowska accordingly illustrates the improvement of, among others, the definition of the target *genius* so that its defining vocabulary matches that of its source.

A fourth issue is that of the absence of reference to the source in the *definiens* which, though scarce, happens with *casserole* in COBUILD4 and OALDCE7. In this case, only reference to the 'food' sense of the word, i.e. the target of the metonymy, and not to the 'recipient' sense, i.e. the source, is made explicitly by those dictionaries. Similarly questionable is what Wojciechowska reports as the absence of reference to the target in the source definition, as we see in the target definition of *press* in LDOCE4. '[P]eople who write reports for newspapers, radio or television' captures the 'reporters' sense, i.e. the target, of the metonymic pattern in the source definition. To our mind, however, it seems a case of absence of reference to the source, thus belonging to the subsection 3.5.4. Instead of the 'mass media' sense, i.e. the source, we only find reference to the 'reporters' sense, i.e. the target. These two issues of inclusion seem to be lexeme dependent. The final problem in definitional terms is that of the misconceived direction of the transfer, more prone to appear with the *definiendum* or anaphoric reference strategies.

In chapter 4 (26 pages), the distinction between count and mass nouns is acknowledged by the author as one of the most problematic for English language learners. Three aspects influence the intelligibility of such aspects in MLDs. Thanks to the trend of pedagogical lexicography, coding has been made more easily understandable for students through the letters [C] for countable and [U] for uncountable. The extra column that the COBUILD4 dictionary has is well-known for its rich grammatical information. The evolution of grammatical coding is also reflected in the inclusion of examples of three types, namely invented, extracted from corpora and corpora-based ones. Whereas examples are mostly used by beginners, more advanced learners resort to pattern illustrations such as *reward sb with sth*. Additionally, definitions are seen as 'much less obvious sources of grammatical information than examples' (2012: 142) by Bogaards and Van der Kloot (2002) and Dziemianko (2006). Wojciechowska favours well-formulated examples in her study as a clarifying strategy for count–mass distinction.

The author emphasizes in subsection 4.2 the crucial role that context plays in making sense of metonymies. She provides the taxonomies in Panther and Thornburg (2003, 2005, 2007) and Paradis (2004). In the latter, for instance, *The pasta bake asked for some more wine*, likely to be uttered in a restaurant setting, relies more heavily on context than *There are a lot of good heads at our university* does, which makes reference to heads as an important part of our human bodies (Paradis 2004: 7). The last empirical study of the dissertation identifies which metonyms out of those 100 in the first study show a countability dissimilarity and which of those are most helpfully encoded. The grammatical coding of 40 metonyms is the target for analysis. Quantitatively, CALD2 presents the worst results assigning confusing or erroneous [C] and [U] in 20 out of 40 instances whereas MEDAL2 assigns countability coding properly in all instances. In terms of examples which were rendered extremely useful for noting this distinction, COBUILD4 achieves the highest results due to its policy of maximum exemplification (Fox 1987: 137).

Although satisfactory overall (except for CALD2 results), four issues are raised by the author concerning the count–mass distinction in dictionaries, namely their occasional opaqueness, discrepancies, inclusion in definitions and absence of exemplification. As the author reports, the [C]/[U] distinctions run the risk of being opaque when the metonymic source and target are joined by a semicolon or the conjunction *or*. In this respect, 'the division of senses into subsenses and providing each with a code is a successful policy' (2012: 152). As far as the second is concerned, 15 out of 40 metonyms lack either [C] or [U] in one of its senses. For instance, the source *beer* in the CONTENTS–CONTAINER pattern refers to the uncountable noun 'beer' as in *I don't drink beer, only spirits*, and to the countable noun 'beer' referring to 'kinds of beer', as in *We have quite a good range of beers*. However, only two (COBUILD4 and OALDCE7) of the five dictionaries under analysis acknowledge that double countability coding of the source *beer*. The last issue, that of examples in the entries of the dictionary, evinces that there is no correlation between the frequency of words and their likelihood of exemplification. Surprisingly, the author notes that *baseball*, *basketball* and *football*, quite common words and part of the EVENT–INSTRUMENT pattern, are not exemplified in any of the 'Big Five'.

On the whole, the general findings of Wojciechowska (2012: 98) regarding the treatment of metonymy in those dictionaries yield diverging results. Although small inconsistencies in dictionaries are present both internally and in relation to others, the Macmillan dictionary outperforms the others regarding cognitive compatibility due to a conscious approach to figurative meaning. For this dictionary, inclusion of 'metonymy boxes' in the fashion of its 'metaphor boxes' is advised by the author. Three techniques are found to be used in the *definiens* of conventionalized metonyms, namely the use of anaphora, the repetition of the *definiendum* or the use of controlled defining vocabulary. The main problems with definitions are that some of them seem to suggest that the source is derived from the target and also seem to obscure the derivation of the

metonymic target from the core meaning. In the final chapter, it is reported that labels indicating countability of nouns of the *beer*, *shipment* and *fleece* type may sometimes be confusing for the learners. Their main feature should be clarity as to whether the source is countable, uncountable or both. According to Wojciechowska (2012), each subsense should have its own example for learners not to confuse the noun's countability. In this respect, examples taken as bearing, among other functions, that of reinforcing grammatical information may help clarify what is countable and uncountable in the metonymic relation.

In Wojciechowska's (2012) understanding, metonymy acts as a process of semantic elaboration of the source into the target and as such, the former should precede the latter. She proposes, though not uncritically, an organization of entries according to Van der Meer's (1997, 1999) and Geeraerts' (1990) ideas. Thence, the core, i.e. prototypical, meaning should come first, and figuratively derived ones should be listed as its subsenses. Metonymy as a same-domain mapping is thought of as closer to the core meaning and is therefore placed before the different-domain mapping operation of metaphor. The author acknowledges the existence of exceptional entries in learner's dictionaries where the core meaning may not come first as with the original meaning of *tea* as the plant: its low frequency may not facilitate the learner's understanding of the word's most common meaning (2012: 31-32). 'Multiple metonymies' as in *tea* or *silver* are accounted for by placing the targets under their source. If they are serially derived, frequency of occurrence will then be used as the criteria to order the metonymic targets. The dictionary entry for a word will thus present the core sense first, then the metonymic source and target, and finally its metaphorical senses.

It must be noted that the author's findings mostly acknowledge the (un)conscious cognitive approach of lexicographers to dictionary making. Some room for improvement is evinced in issues such as the coding of countability in each of the subsenses or consistent use of a defining strategy for metonymies. Nonetheless, MLDs are above all a product designed for the success of language learners. The subtleties and ameliorations proposed by Wojciechowska (2012) must be collated with empirical studies that verify the positive effects of her proposals on L2 learning. Tono (2001: 12-13) supports this claim by acknowledging the gap between the lexicographers' expectations of the users' skills and the actual skills of students in using the dictionary. Since that gap may hinder the entrenchment and apprehension of these metonymic patterns by English language learners, further experimental studies should be conducted to prove the present suggestions successful.

Wojciechowska's (2012) contribution is a step further in the optimization of lexicographic works regarding current semantic theories with an emphasis on metonymy codification. Her book presents the state of affairs of some conventionalized metonymic patterns as reflected in the 'Big Five'. Her study endorses the principle that '[e]xplanations of word meanings should rather be as intelligible as possible even if it sometimes means sacrificing a dictionary's uniformity' (2012: 119). Despite the naturally limited though rich analysis, the

author mentions but does not model 'template entries' (Atkins and Rundell 2008) for the standardization of dictionaries' microstructure of metonyms. On the whole, her research as presented in *Conceptual Metonymy and Lexicographic Representation* (2012) adds on the pedagogical lexicographic trends which advocate a more transparent connection between literal and figurative meaning.

References

- Adamska-Salaciak, A. 2006. *Meaning and the Bilingual Dictionary. The Case of English and Polish*. Frankfurt am Main: Peter Lang Verlag.
- Apresjan, J.D. 1995. *Semantyka Leksykalna. Synonimiczne Środki Języka*. Wrocław: Zakład Narodowy im. Ossolińskich-Wydawnictwo.
- Atkins, B.T.S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Barcelona, A. 2000. Introduction: The Cognitive Theory of Metaphor and Metonymy. Barcelona, A. (Ed.). 2000. *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*: 1-25. Berlin/New York: Mouton de Gruyter.
- Boers, F. 2000. Metaphor Awareness and Vocabulary Retention. *Applied Linguistics* 21(4): 553-571. Available: <http://sla.sjtu.edu.cn/thesis/Metaphor%20awareness%20and%20vocabulary%20retention.pdf>.
- Boers, F. and S. Lindstromberg. 2006. Cognitive Linguistic Applications in Second or Foreign Language Instruction: Rationale, Proposals and Evaluation. Kristiansen, G., M. Achard, R. Dirven and F.J. Ruiz de Mendoza Ibáñez (Eds.). 2006. *Cognitive Linguistics: Current Applications and Future Perspectives*: 305-358. Berlin/New York: Mouton de Gruyter.
- Bogaards, P. and W.A. van der Kloot. 2002. Verb Constructions in Learners' Dictionaries. Braasch, A. and C. Povlsen. 2002. *Proceedings of the Tenth EURALEX International Congress. EURALEX 2002. Volume II: 747-757*. Copenhagen: Center for Sprogteknologi.
- Croft, W. 2002. The Role of Domains in the Interpretation of Metaphors and Metonymies. Dirven, R. and R. Pörings (Eds.). 2002. *Metaphor and Metonymy in Comparison and Contrast*: 161-205. Berlin/New York: Mouton de Gruyter.
- Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- DeCesaris, J. and V. Alsina. 2002. The Representation of Figurative Senses in Learner's Dictionaries. Gottlieb, H., J.E. Mogensen and A. Zettersen (Eds.). 2002. *Symposium on Lexicography X: Proceedings of the Tenth International Symposium on Lexicography May 4-6, 2000 at the University of Copenhagen*: 77-90. Tübingen: Max Niemeyer.
- Dziemiątko, A. 2006. *User-friendliness of Verb Syntax in Pedagogical Dictionaries of English*. Lexicographica. Series Maior 130. Tübingen: Max Niemeyer.
- Fox, G. 1987. The Case for Examples. Sinclair, J.M. (Ed.). 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*: 137-149. London: Collins ELT.
- Geeraerts, D. 1990. The Lexicographical Treatment of Prototypical Polysemy. Tsohatzidis, S.L. (Ed.). 1990. *Meaning and Prototypes: Studies in Linguistic Categorisation*: 195-210. London: Routledge.
- Hanks, P. 2000. Do Word Meanings Exist? *Computers and the Humanities* 34: 205-215.

- Kilgarriff, A.** 1997. I Don't Believe in Word Senses. *Computers and the Humanities* 31(2): 91-113.
- Lakoff, G.** 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Lakoff, G. and M. Johnson.** 1980. *Metaphors We Live By*. Chicago/London: University of Chicago Press.
- Lakoff, G. and M. Turner.** 1989. *More than Cool Reason. A Field Guide to Poetic Metaphor*. Chicago/London: The University of Chicago Press.
- Landau, S.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Cambridge/New York: Cambridge University Press.
- Langacker, R.W.** 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Mel'čuk, I.** 1988. Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria. *International Journal of Lexicography* 1(3): 165-188.
- Nerlich, B. and D.D. Clark.** 2001. Serial Metonymy: A Reference-based Study of Polysemisation. *Journal of Historical Pragmatics* 2(2): 245-272.
- Panther, K.-U. and G. Radden (Eds.).** 1999. *Metonymy in Language and Thought*. Amsterdam/Philadelphia: John Benjamins.
- Panther, K.-U. and L. Thornburg.** 2003. Metonymies as Natural Inference and Activation Schemas: The Case of Dependent Clauses as Independent Speech Acts. Panther, K.-U. and L. Thornburg (Eds.). 2003. *Metonymy and Pragmatic Inferencing*: 127-147. Amsterdam/Philadelphia: John Benjamins.
- Panther, K.-U. and L. Thornburg.** 2005. Inference in the Construction of Meaning: The Role of Conceptual Metonymy. Górska, E. and G. Radden (Eds.). 2005. *Metonymy–Metaphor Collage*: 37-57. Warsaw: Warsaw University Press.
- Panther, K.-U. and L. Thornburg.** 2007. Metonymy. Geeraerts, D. and H. Cuyckens (Eds.). 2007. *The Oxford Handbook of Cognitive Linguistics*: 236-263. Oxford/New York: Oxford University Press.
- Paradis, C.** 2004. Where Does Metonymy Stop? Senses, Facets and Active Zones. *Metaphor and Symbol* 19(4): 245-264. Available: <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=135358&fileId=1590153>.
- Peirsman, Y. and D. Geeraerts.** 2006. Don't Let Metonymy Be Misunderstood. An Answer to Croft. *Cognitive Linguistics* 17(3): 269-316.
- Philip, G.** 2006. Metaphor, the Dictionary, and the Advanced Learner. Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings of the Twelfth EURALEX International Congress, Torino, Italia, September 6th–9th, 2006. Volume 2*: 895-905. Alessandria: Edizioni dell'Orso. Available: <http://amsacta.unibo.it/2062/1/MetaphorDictionary.pdf>.
- Radden, G. and Z. Kövecses.** 1999. Towards a Theory of Metonymy. Panther, K.-U. and G. Radden (Eds.). 1999: 17-59.
- Ruiz de Mendoza Ibáñez, J.F.** 2000. The Role of Mappings and Domains in Understanding Metonymy. Barcelona, A. (Ed.) 2000. *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*: 109-132. Berlin/New York: Mouton de Gruyter.
- Seto, K.** 1999. Distinguishing Metonymy from Synecdoche. Panther, K.-U. and G. Radden (Eds.). 1999: 91-120.

- Tono, Y.** 2001. *Research on Dictionary Use in the Context of Foreign Language Teaching: Focus on Reading Comprehension*. Lexicographica. Series Maior 106. Tübingen: Max Niemeyer.
- Ungerer, F. and H.-J. Schmid.** 1996. *An Introduction to Cognitive Linguistics*. London/New York: Longman.
- Van der Meer, G.** 1997. Four English Learner's Dictionaries and their Treatment of Figurative Meanings. *English Studies* 78(6): 556-571.
- Van der Meer, G.** 1999. Metaphors and Dictionaries: The Morass of Meaning, or How to Get Two Ideas for One. *International Journal of Lexicography* 12(3): 195-208.
- Van der Meer, G. and R. Sansome.** 2001. OALD6 in a Linguistic and a Language Teaching Perspective. *International Journal of Lexicography* 14(4): 283-306.
- Warren, B.** 1999. Aspects of Referential Metonymy. Panther, K.-U. and G. Radden (Eds.). 1999: 121-138.
- Wojciechowska, S.** 2010. Metonymy Representation in English Monolingual Learners' Dictionaries: Problems and Solutions. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the 14th EURALEX International Congress, Leeuwarden/Ljouwert, The Netherlands, 6-10 July 2010*: 1436-1449. Ljouwert: Afûk. Available: http://www.euralex.org/elx_proceedings/Euralex2010/136_Euralex_2010_9_WOJCIECHOWSKA_Metonymy%20representation%20in%20English%20monolingual%20learners%20dictionaries_Problems%20and%20solut.pdf.
- Wojciechowska, S.** 2012. *Conceptual Metonymy and Lexicographic Representation*. Frankfurt am Main: Peter Lang.

José Antonio Jódar Sánchez
San José State University
San José, California
United States of America
(jjodarsa7@alumnes.ub.edu)

Publikasieaankondigings / Publication Announcements

W.F. Botha (Hoofredakteur). *Woordeboek van die Afrikaanse Taal, Veertiende Deel: S-SKOOI*. 2013, xxi + 674 pp. ISBN-13 978-0-9814434-6-1 (leerband), ISBN-13 978-0-9814434-5-4 (plastiekband). Stellenbosch: Buro van die WAT. Prys: R650 (leerband) / R350 (plastiekband). (Resensieartikel in hierdie nommer.)

Julie Coleman. *A History of Cant and Slang Dictionaries, Volume IV: 1937–1984*. 2010, xxvi + 503 pp. ISBN 978 0 19 956725 6 (Hardback). Oxford: Oxford University Press. Price: £85. (Review in this issue.)

G.-M. de Schryver et al. (Editors). *Oxford Bilingual School Dictionary: IsiXhosa and English*. 2014, 562 pp. ISBN 978-0-19-576682-0. Cape Town: Oxford University Press Southern Africa. Price R129.95. (Review in this issue.)

Pedro A. Fuertes-Olivera and Sven Tarp. *Theory and Practice of Specialised Online Dictionaries*. 2014, x + 272 pp. ISBN 978-3-11-034883-5, e-ISBN 978-3-11-034902-3. Berlin/New York: Mouton de Gruyter. Price: €99.95. (Review article in this issue.)

Mariusz Kamiński. *A History of the Chambers Dictionary*. 2013, xii + 297 pp. ISBN 978-3-11-031250-8, e-ISBN 978-3-11-031273-7, ISSN 0175-9264. Lexicographica. Series Maior 143. Berlin/Boston: Walter de Gruyter. Price: €84. (Review in this issue.)

Taasisi ya Taaluma za Kiswahili (TATAKI). *Kamusi ya Kiswahili Sanifu*. Third Edition. 2013, xvi + 687 pp. ISBN 978-019-573820-9. Nairobi: Oxford University Press. Price: Kshs. 840.00 (Kenyan Shilling). (Review in this issue.)

Sylwia Wojciechowska. *Conceptual Metonymy and Lexicographic Representation*. 2012, 209 pp. ISBN 978-3-631-63510-0. Frankfurt am Main: Peter Lang. Price: €53.15. (Review in this issue.)

VOORSKRIFTE AAN SKRYWERS

(Tree asseblief met ons in verbinding (lexikos@sun.ac.za) vir 'n uitvoeriger weergawe van hierdie instruksies of besoek ons webblad: <http://www.wat.co.za>)

A. REDAKSIONELE BELEID

1. Aard en inhoud van artikels

Artikels kan handel oor die suiwer leksikografie of oor implikasies wat aanverwante terreine, bv. linguistiek, algemene taalwetenskap, terminologie, rekenaarwetenskap en bestuurskunde vir die leksikografie het.

Bydraes kan onder ingeen 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 evaluerend 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) **Ander:** Van tyd tot tyd kan ander rubrieke deur die redaksie ingevoeg word, soos Leksikoprogrammatuur, Leksiko-opname, Leksikobibliografie, Leksikonuus, Lexikofokus, Leksiko-eerbewys, Leksikohuldeblyk, 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 Hoër Onderwys van die Suid-Afrikaanse Regering as 'n gesubsidieerde, d.w.s. inkomstegenererende navorsingstydskrif goedgekeur. Dit verskyn ook op die *Institute of Science Index (ISI)*.

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.

Manuskripte word vir publikasie oorweeg met dien verstande dat die redaksie die reg voorbehou om veranderinge aan te bring om die styl en aanbieding in ooreenstemming met die redaksionele beleid te bring. Outeurs moet toesien dat hulle bydraes taalkundig en stilisties geredigeer word voordat dit ingelewer word.

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* gepubliseer 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. Indien van toepassing, moet besonderhede van die oorsprong van die artikel (byvoorbeeld 'n referaat by 'n kongres) verskaf word.

6. Gratis oordrukke en eksemplare

Skrywers ontvang vyf gratis oordrukke van elke artikel van hulle wat gepubliseer is asook een gratis eksemplaar van die uitgawe waarin sodanige artikel(s) verskyn het. Skrywers van suiwer evaluerende resensies en van bydraes tot die rubrieke Leksikonotas, Leksikovaria, ens. ontvang slegs vyf gratis oordrukke van hulle bydraes. In laasgenoemde 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 verkieslik in elektroniese formaat aan die volgende adres te stuur: lexikos@sun.ac.za, of Die Redakteur: LEXIKOS, Buro van die WAT, Postbus 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

Manuskrip moet verkieslik in elektroniese formaat per e-pos of op rekenaarskyf voorgelê word in sagteware wat versoenbaar is met MS Word. Die lettersoort moet verkieslik 10-punt Palatino of Times Roman wees. Bydraes moet verkieslik nie 8 000 woorde oorskry nie.

Elke artikel moet voorsien wees van 'n opsomming van ongeveer 200 woorde en ongeveer 10 sleutelwoorde in die taal waarin dit geskryf is, sowel as 'n opsomming en sleutelwoorde in Engels. Engelse artikels van Suid-Afrikaanse oorsprong moet 'n opsomming en sleutelwoorde in Afrikaans hê, terwyl Engelse artikels van buitelandse oorsprong 'n tweede opsomming en sleutelwoorde in enigen van die aangeduide tale mag gee. As die outeur dit nie doen nie, sal die redaksie 'n Afrikaanse vertaling voorsien. Maak seker dat die opsomming in die tweede taal ook 'n vertaling van die oorspronklike titel bevat.

2. Grafika

Figure, soos tabelle, grafieke, diagramme en illustrasies, moet in 'n gepaste grootte wees dat dit versoen kan word met die bladspieël van *Lexikos*, naamlik 18 cm hoog by 12 cm breed. Die plasing van grafika binne die teks moet duidelik aangedui word. Indien skryftekens of grafika probleme oplewer, mag 'n uitdruk van die manuskrip of 'n e-pos in .pdf-formaat aangevra word.

3. Bibliografiese gegewens en verwysings binne die teks

Kyk na onlangse nommers van *Lexikos* vir meer inligting.

4. Aantekeninge/voetnote/eindnote

Aantekeninge moet deurlpend in die vorm van boskrite genommer en aan die einde van die manuskrip onder die opskrif **Eindnote** gelys word.

INSTRUCTIONS TO AUTHORS

(For a more detailed version of these instructions, please contact us (lexikos@sun.ac.za) or refer to our website: <http://www.wat.co.za>)

A. EDITORIAL POLICY

1. Type and content of articles

Articles may treat pure lexicography or the implications that related fields such as linguistics, general linguistics, terminology, computer science and management have for lexicography.

Contributions may be classified in any one of the following categories:

(1) **Articles:** Fundamentally original scientific research done and the results obtained, or existing research results and other facts reflected 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 containing announcements and press releases by lexicographic societies which are of particular value to the practising lexicographer.

(7) **Other:** From time to time other categories may be inserted by the editors, such as Lexicosoftware, Lexicosurvey, Lexicobibliography, Lexiconews, Lexicofocus, Lexicohonour, Lexicotribute, 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 Higher Education of the South African Government has approved *Lexikos* as a subsidized, i.e. income-generating research journal. It is also included in the *Institute of Science Index (ISI)*.

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.

Manuscripts are considered for publication on the understanding that the editors reserve the right to effect changes to the style and presentation in conformance with editorial policy. Authors are responsible for the linguistic and stylistic editing of their contributions prior their submission.

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. If applicable, details about the origin of the article (e.g. paper read at a conference) should be supplied.

6. Free offprints and copies

Authors will receive five free offprints of each of their 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, etc., receive five free offprints of their contributions. In the case of the latter categories, the editors may, 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, preferably in electronic format, for publication in *Lexikos* to: lexikos@sun.ac.za, or

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. Format and length of articles

Manuscript should preferably be submitted in electronic format by email or on a disk, in software compatible with MS Word. The typeface used should preferably be 10-point Palatino or Times Roman. Contributions should not exceed **8 000 words**.

Each article must be accompanied by **abstracts** of approximately 200 words and approximately **10 keywords** in the language in which it is written, as well as **in English**. English articles of South African origin should carry an abstract and keywords in Afrikaans, whilst English articles of foreign origin should carry a second abstract and keywords in any of the other languages mentioned. In cases where this is not done, the editors will provide an Afrikaans version. Ensure that the abstract in the second language also contains a **translation of the original title**.

2. Graphics

Figures such as tables, graphs, diagrams and illustrations should be in an appropriate size to be well accommodated within the page size of *Lexikos*, namely 18 cm high by 12 cm wide. The locations of figures within the text must be clearly indicated. If orthographic marks or graphics used in the text prove problematic, a printout of the manuscript or an email in .pdf format may be requested.

3. Bibliographical details and references in the text

Examine recent issues of *Lexikos* for details.

4. Notes/footnotes/endnotes

Notes must be numbered consecutively by superscript numbers and grouped together at the end of the manuscript under the heading **Endnotes**.